SCIENTIFIC AMERICAN

FEBRUARY 1993

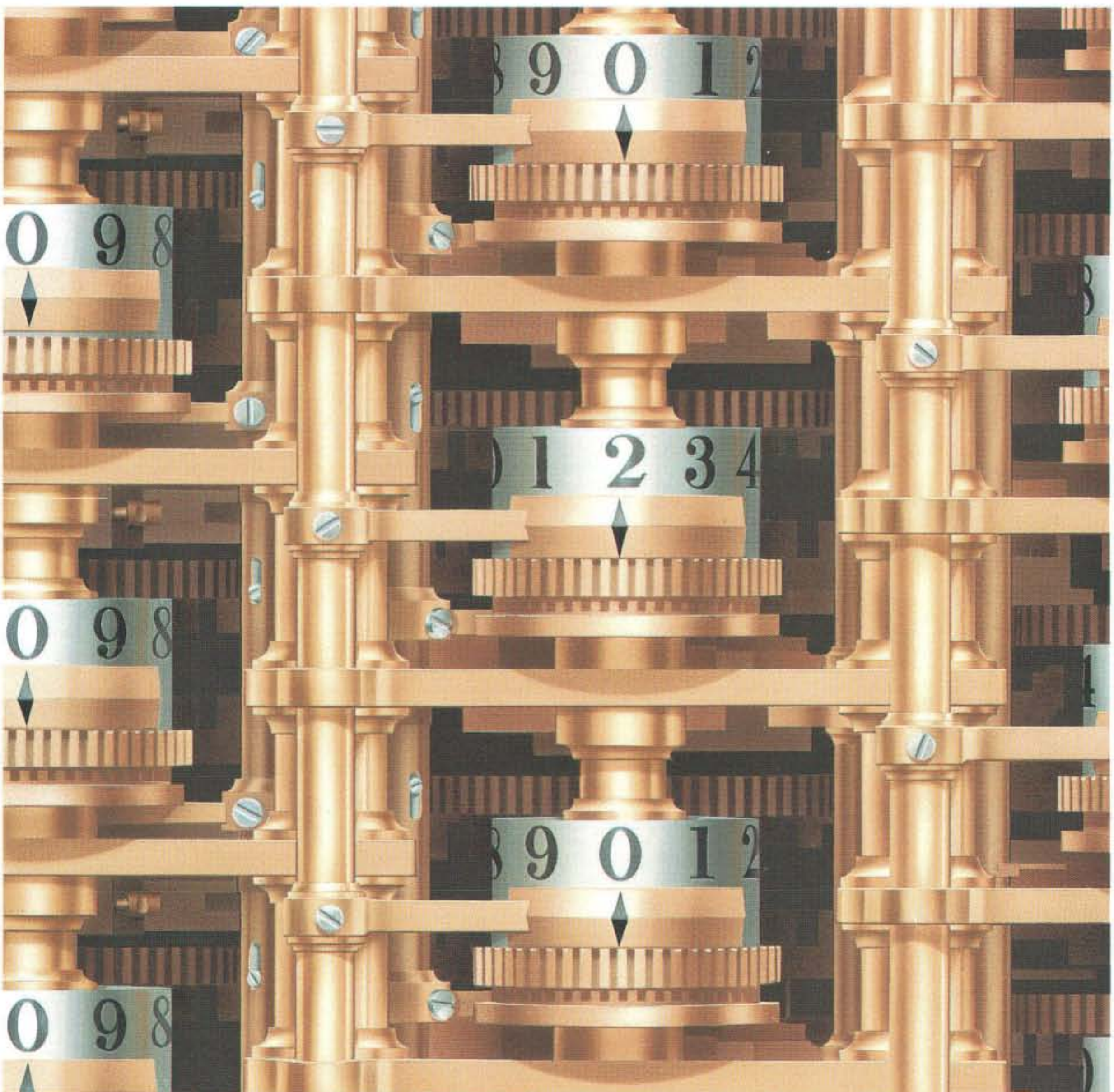
$3.95
U.K. £2.50

RUSSIAN SCIENCE IN CRISIS

Beating resistance in superconductors.

Violence and the environment.

Zinc fingers that help switch on genes.



02

9 770036 873015

***A calculating engine** was built more than a century after it was attempted by Charles Babbage. It works.*

To your family, to your business associates, to a potential client, there is nothing more powerful than the sound of your voice.

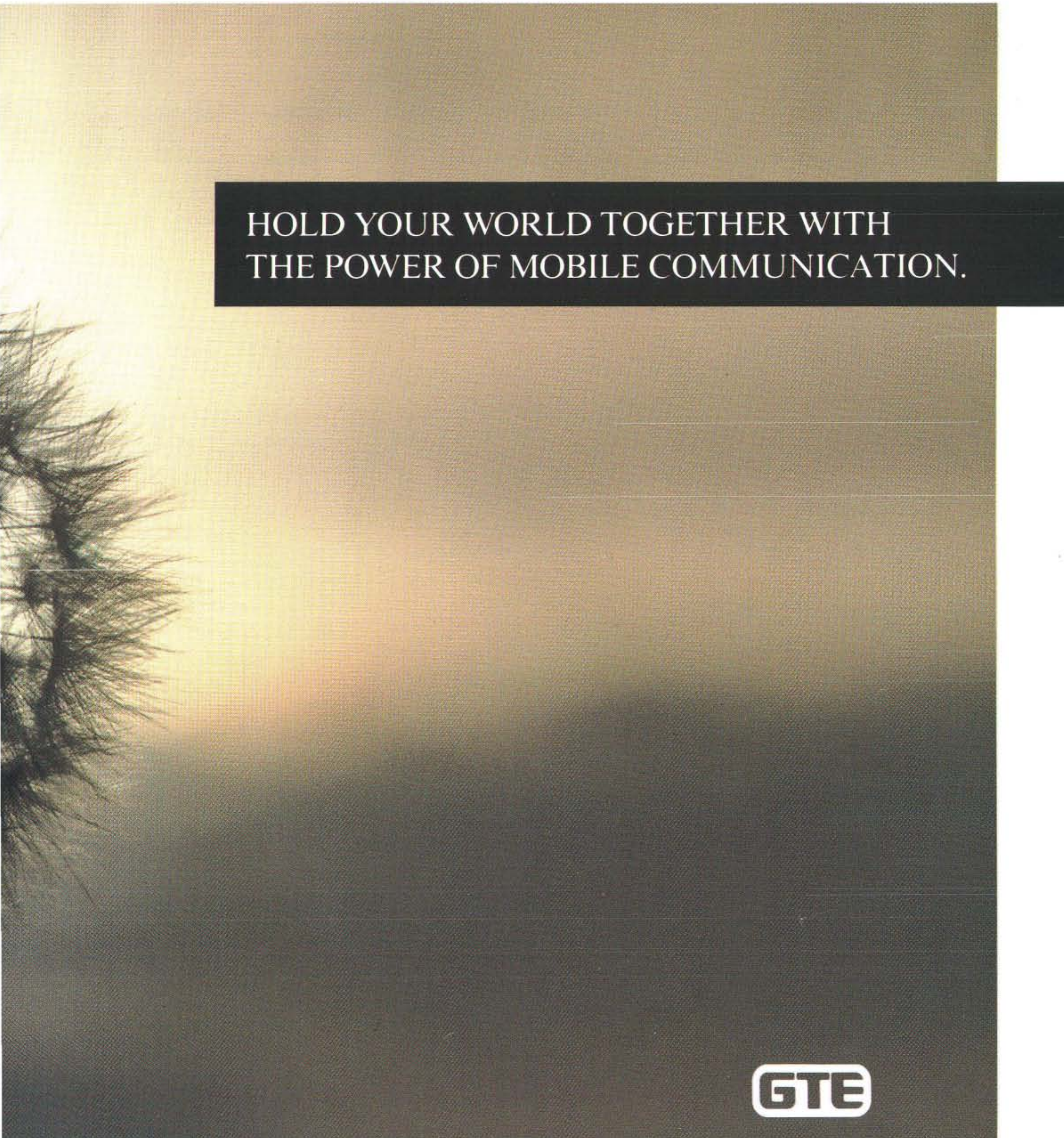
At GTE, we can help you keep that power in the palm of your hand. To take with you beyond new hor zons, beyond all past limitations.

With GTE's growing Mobilnet® system, for example, your car telephone keeps you in touch in over 28

metropolitan areas. From Tampa to San Francisco to the state of Hawaii. GTE's Airfone® In-flight Telephone Service is now on 15 major airlines. And Amtrak is expanding the availability of Railfone®.

Offer a comforting word to your child, check in with your office, open the door of a new business opportunity. At GTE, we give you the power to touch your world.

Because at GTE, the power is on.



SCIENTIFIC AMERICAN

February 1993 Volume 268 Number 2

16

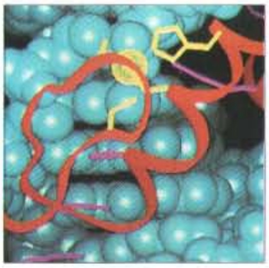

Environmental Change and Violent Conflict
Thomas F. Homer-Dixon, Jeffrey H. Boutwell and George W. Rathjens

It has long been predicted that a collision between a growing world population and increasing environmental degradation would lead to civil and international strife. A team of researchers commissioned to study the evidence believes that day may have arrived. Shortages of water, forests and fertile land are already contributing to violent conflicts in many parts of the developing world.

24


Resistance in High-Temperature Superconductors
David J. Bishop, Peter L. Gammel and David A. Huse

The discovery that certain ceramics conduct electricity with no resistance at comparatively balmy temperatures had researchers eyeing a range of applications. But the materials quickly betrayed a critical flaw: in a magnetic field, they lose their ability to superconduct. The mechanism of resistance is now understood, raising the prospect that the problem can be controlled.

32


Zinc Fingers
Daniela Rhodes and Aaron Klug

These projections on transcription factors grip specific sites on DNA, preparing genes for activation. Since they were discovered in 1985, proteins incorporating zinc fingers have been identified in diverse species, from yeast to humans. Several laboratories have begun to decipher how these zinc-containing proteins select and bind to DNA and to elucidate the role they play in switching on genes.

40


How Should Chemists Think?
Roald Hoffmann

The molecules that exist naturally on the earth and those made in laboratories are produced by a common process: synthesis. When chemists design new compounds, they can either emulate nature or be guided by the whims of the mind. The author explores the paradoxes that arise by describing the creation of a widely used antibiotic and an utterly useless, perfectly beautiful iron compound.

50

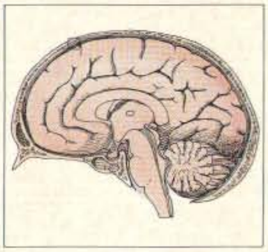

SCIENCE IN PICTURES

A Technology of Kinetic Art
George Rickey

This sculptor's dynamic works seem to balance uncannily and stir in the slightest gust of wind. The laws of physics that govern pendulums serve as the foundation of this intricate choreography of weight and balance.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 1993 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111, or fax: (212) 355-0408.

56



Breaching the Blood-Brain Barrier
Elaine Tuomanen

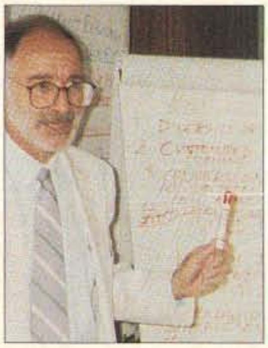
The blood-brain barrier is not so impervious as it seems. Some bacteria, especially those that cause meningitis, manage to sneak across. By developing a treatment for this fatal disease, the author has discovered clues to the process that may allow physicians to smuggle drugs into the brain for treating tumors and other disorders.

62

Redeeming Charles Babbage's Mechanical Computer
Doron D. Swade

Historians have argued that Charles Babbage was unable to build his vast mechanical computers because his conception exceeded the capacity of 19th-century engineering. The construction in 1991 of a working, three-ton calculating engine proves that his designs were well within the realm of possibility.

68



TRENDS IN RUSSIAN SCIENCE

Selling to Survive
Tim Beardsley, staff writer

The disintegration of the Soviet Union and the near collapse of its scientific institutions have plunged researchers into a battle for their professional lives. Many have left; others have put their talent up for sale. Western corporations have found a buyer's market of research capability. But will such efforts tide the Russian scientific establishment over the disruption of economic and political reform?

DEPARTMENTS

7

Science and the Citizen

The odds of finding habitable planets.... Escher and Penrose.... Corroborating *COBE*.... The controversy over genes and crime.... Are submarines still undetectable?... PROFILE: Nathan P. Myhrvold, Microsoft's advanced technology wizard.

78

Science and Business

Automated eyes for the Postal Service.... Making learning part of the job.... Fending off lightning bolts.... Nutty ideas.... Teaching physics with virtual reality.... THE ANALYTICAL ECONOMIST: Why industry leaders are not nimble innovators.

5

Letters to the Editors
Taxonomic conundrum.... Why more women are not engineers.

6

50 and 100 Years Ago
1893: Skeletal evidence for walking in circles when lost.

85

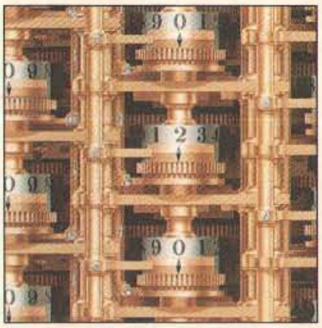
Mathematical Recreations
The fuzzy logic between being totally true and totally false.

88

Book Review
How to preserve the planet when human activity is a major force.

92

Essay: *David C. Cassidy*
The real reason Germany lost the race to build the atomic bomb.



THE COVER painting shows a detail of the partially completed Difference Engine No. 1, an automatic calculator designed by Charles Babbage in the 1820s. Baggage's plans for mechanical calculators and computers paved the way for the modern computer revolution, but he never managed to build any of his devices in its entirety. A recent reconstruction of one of his calculators proves that his designs were in fact logically sound and practically feasible [see "Redeeming Charles Babbage's Mechanical Computer," by Doron D. Swade, page 62].

THE ILLUSTRATIONS

Cover painting by George Retseck

Page	Source
17	Baldev/Sygma
18	Jared Schneidman/JSD
19	Jared Schneidman/JSD (*top*), Thomas F. Homer-Dixon (*bottom*)
20–21	Jared Schneidman/JSD
22	Jared Schneidman/JSD (*top*), Mike Goldwater–Network/Matrix (*bottom*)
23	Jared Schneidman/JSD (*left*), Thomas F. Homer-Dixon (*right*)
25	David J. Bishop and Carlos A. Durán, AT&T Bell Laboratories
26	Jared Schneidman/JSD
27	Jared Schneidman/JSD (*top*), David J. Bishop (*middle* and *bottom*)
28	Jared Schneidman/JSD
29	Jared Schneidman/JSD (*left*), David J. Bishop (*right*)
30	Jared Schneidman/JSD (*left*), Robert Prochnow (*right*)
31	Jared Schneidman/JSD
33	Kirk Moldoff
34	Guilbert Gates/JSD
35	David Neuhaus, Medical Research Council (MRC) Laboratory of Molecular Biology (*left*), Daniela Rhodes, MRC (*right*)
37	Guilbert Gates/JSD (*top*), Gabor Kiss (*bottom*)
38	Guilbert Gates/JSD
39	John W. R. Schwabe and Daniela Rhodes, MRC
41	Art Resource, Inc.
42–43	Boris Starosta; Ralph Mosley/Merck & Co. (*insets*)
44–45	Kingsley L. Taft, Massachusetts Institute of Technology
48–49	Boris Starosta
50–53	Christopher Burke, Quesada/Burke; Jana Brenning (*drawings*)
54	Achim Pahle (*top*), Peter Hollenbach, Christoph & Mayer (*bottom*)
55	Christopher Burke, Quesada/Burke (*left* and *bottom*), TGL, Williamstown, Mass. (*top right*)
56–57	Carol Donner
58	Ian Worpole (*top*), Elaine Tuomanen (*bottom*)
59	Carol Donner
60	Yoav Levy/Phototake, Inc. (*left*), Johnny Johnson (*right*)
62–65	Doron D. Swade (photographs by David Exton/Science Museum Photostudio)
66	Johnny Johnson
67	Doron D. Swade (photographs by David Exton/Science Museum Photostudio)
68–69	sovfoto/eastfoto
70–72	Andrew P. Amelin
73	Courtesy of Boris V. Kuteev
74	Andrew P. Amelin
75	Tim Beardsley
76	Robert Summers
85	Patrick Grim (*left*), Jared Schneidman/JSD (*right*)

SCIENTIFIC AMERICAN®

Established 1845

LETTERS TO THE EDITORS

Count on Confusion

Robert M. May makes excellent points in "How Many Species Inhabit the Earth?" [SCIENTIFIC AMERICAN, October 1992]. I was especially taken by his suggestion that butterflies have attained the "honorary status of birds." Giving the currently known species of butterflies as 17,500, he estimates the true number as no more than 20,000. Later in the same issue ("Singing Caterpillars, Ants and Symbiosis"), Philip J. DeVries cites the number of known butterfly species as "more than 13,500." It presents a nearly perfect example of May's central thesis concerning the uncertainty of the number of taxa.

CHARLES E. DITERS
U.S. Fish and Wildlife Service

Sex Ratios at Work

I am concerned that some of the opinions in "Sex Differences in the Brain," by Doreen Kimura [SCIENTIFIC AMERICAN, September 1992], are misleading and potentially damaging. Your readers deserve to know that Kimura's opinion regarding a biological foundation for occupational sex segregation is not shared by all scientists.

Whether the measured sex differences in certain cognitive and motor skills are "quite substantial" as she says is debatable. Certainly, none of them develops independent of social influences. Even if they did, the ratio of men and women in science and engineering would be closer to 50/50. In some fields of science and engineering, the current sex ratio is more than 90 percent men to fewer than 10 percent women.

Kimura indicates that the sex differences range from approximately 0.20 standard deviation for one measure of verbal fluency to approximately 0.75 standard deviation for one of targeting skill. She calls the 0.75 effect size large. Yet the sex difference in adult height in the U.S. is approximately 2.0 standard deviations. Thus, even the largest sex difference on any individual cognitive or motor test is substantially smaller than the sex difference in height. The largest sex difference on any ability construct (defined by performance on several related tests) is that in visuospatial ability, which is only about 0.45 effect size units—a little less than one quarter the difference in height.

Using an extreme assumption that visuospatial ability is the only factor determining success as an engineer or physicist, one would expect about 60 percent of those jobs to be held by men and about 40 percent by women. Even if a person needed to score in the top 5 percent of the population in visuospatial ability to succeed, a ratio of only about 70 men to 30 women would be predicted. Those predictions assume that the sex difference is determined exclusively by factors that cannot be modified by socialization or education, which is not true.

Researchers studying sex segregation in occupations have concluded that the major determinants are economic and political, not hormonal. It would be difficult to explain the major shifts in fields such as teaching and secretarial work, which men once dominated, in terms of biology. If women continue to be misinformed about their chances of succeeding as engineers and scientists, the sex ratios in those professions are unlikely to change. As Bernadine Healy, the director of the National Institutes of Health, stated in 1991, "It is safe to say that sustaining America's scientific preeminence will depend on attracting—and retaining—talented women." Perpetuation of stereotypes about sex and science works against this goal.

MELISSA HINES
Department of Psychiatry and Biobehavioral Sciences
School of Medicine, University of California, Los Angeles

Kimura replies:

I agree that the reasons men and women are differentially represented across occupations are complex. Nevertheless, my claim that on the basis of biological predisposition, men and women would not be expected to be *equally* represented in all occupations is, I believe, a moderate view shared by most biological scientists in this field (most of whom are women).

If one looks at a specific visuospatial ability such as mental rotation, the differences between men and women range across studies from 0.70 to 1.0 in effect size. The sex differences in mathematical reasoning hover around 0.50. Even in the latter case, the ratio of men to women at the upper end of the distribution is very high, and it is differentiation at the upper end that is significant for certain professions. A recent study reported that girls with very high math achievement scores also tend to have interests and values that better suit them for nonscience fields. Such values are not necessarily determined by socialization.

The common inference that women are kept out of the sciences by systemic or deliberate discrimination is not based on evidence. One might as well argue that men are kept out of nursing careers by discrimination. Instead the process appears to be largely self-selection. As for the desirability of attracting women to the physical sciences, that is a political, not a scientific, issue.

Still Scavenging

Others have observed that modern parks appeal to us by recapitulating the East African savanna of our hominid ancestors. If Robert J. Blumenschine and John A. Cavallo ["Scavenging and Human Evolution," SCIENTIFIC AMERICAN, October 1992] are right, another taste from that time may remain. The meat we buy in the supermarket, though called fresh, has generally been hung for a day or two to "age"—producing exactly the quality our vestigial scavenger instincts still prize: a delicate carrion tang.

STUART GELZER
Ardmore, Pa.

Caveat Educator

Three sample science questions, devised by an individual who casts himself as a reformer of the science curricula in our schools, were posed in "Teaching Real Science," by Tim Beardsley [SCIENTIFIC AMERICAN, October 1992]. It appears that we also need to be concerned about the English curricula, to wit Bill Aldridge's question: "Which coffeepot would hold the most coffee?"

Correct English usage would have been, "Which coffeepot would hold *more* coffee?" One uses the superlative only to compare three or more objects.

E. KENNETH SNYDER
Seattle, Wash.

50 AND 100 YEARS AGO

FEBRUARY 1943

"The government's 'scrap-fat drive' to obtain new sources of glycerin and soap acids asks housewives and restaurants to save their grease drippings for defense. From these scrap fats the government expects to make glycerin for explosives, replacing the imports of cocoanut oil from the Philippines and other Pacific Islands, which were cut off by the war. But there's another not-so-widely publicized source which has been providing the United States with oil for glycerin for two decades—sardines and herring, which inhabit the Pacific Ocean from Alaska to Mexico—and this source has a distinct advantage, in that the product needs only to be harvested; it requires no preliminary planting and cultivation."

"The chances of arresting the development of stuttering are much greater in the primary stage, before anxiety and inferiority feelings begin to develop and before conditioning has had time to operate. Therapy is largely a matter of slowing down the tempo of living and removing any exciting stimuli in the home environment, particularly the excitement and tensions generated by neurotic parents. Family quarrels, exciting games, rapid speech or other 'nervous' reaction patterns on the part of parents or older children should be eliminated. The child should be kept in as good a physical condition as possible, he should have frequent periods of rest and relaxation, and fatigue should be avoided. Also, since the stuttering child demonstrates in general a lowered degree of psychomotor efficiency, especially in those functions requiring fine coordination, a certain amount of rhythmic work is recommended."

"A practically complete skeleton of Barylambda, an extinct mammal which, when it lived 50,000,000 years ago in west-central Colorado, attained a development entitling it to be rated as one of the most heavily built animals of all time, has just been placed on exhibition in the hall of paleontology at the Field Museum of Natural History, Chicago. 'Barylambda was unlike and unrelated to any present-day animal,' states Bryan Patterson, who led the expedition which excavated the remains of the rare creature. 'It stood some four feet high, had an overall length of about eight and a half feet, and its width across the hips was almost equal to three-quarters of its height. Its bones were extraordinarily massive, indicating the possession of immense muscular power.' "

SCIENTIFIC AMERICAN

FEBRUARY 1893

"The fact that people lost on a desert or in a forest invariably walk in a circle is due to slight inequality in the length of the legs. Careful measurements of a series of skeletons have shown that only ten per cent had the lower limbs equal in length, thirty-five per cent had the right limb longer than the left, while in fifty-five per cent the left leg was the longer. The result of one limb being longer than the other will naturally be that a person will unconsciously take a longer step with the longer limb, and consequently will trend to the right or to the left, according as the left or right is the longer, unless the tendency to deviation is corrected by the eye."

"The $3,000,000 which the hat manufacturers of the country have got to hand over to the inventor of the sweat band used on hats affords a striking illustration of the value of genius when it makes a hit."

"**How to Freeze Water on a Small Scale.** Take a concave watch glass, touch the convex side upon water so as to leave a drop hanging from the glass. Pour a little ether into the concave and blow upon it. The rapid evaporation of the ether will render the glass so cold that the drop of water will be frozen."

"The enormous strides made by electricity in commerce and industries have been, to a certain extent, paralleled by applications in medicine and surgery. One of the new features of electric medication is the introduction of drugs into the human body through the skin. This is done by placing solutions of any drug upon a sponge, which is made the positive pole and placed against the skin. When the current is turned on, the drug is actually driven through the skin into the tissues. The application is not at all painful. Thus cocaine has been driven in over a painful nerve, and neuralgias have been relieved by it. Many other drugs have been used in this way. This property of electricity is known as cataphoresis. Operations have been performed after anaesthetizing the skin and subjacent tissues cataphoretically."

"Lion-tigers have been born in several menageries, but the most interesting hybrids seen in the second half of this century were a litter of Siberian fox-dogs [*see illustration at left*]. Their mother had been a spitz and their male progenitor a black fox, and there was, withal, something strangely raccoonish in their appearance that would have warranted the suspicion of a triple *mésalliance* if the *Procyon lotor* were not a total stranger to the fauna of the eastern continent."

Siberian fox-dogs

Livable Planets

Calculations raise the odds for finding life in the cosmos

In the Pink Panther movies, Inspector Clouseau bumbles toward the solution of crimes while remaining untouched by a maelstrom of disasters and mishaps. Life on the earth seems to have navigated a similarly fortunate course. The planet orbits comfortably between hellishly hot Venus and frozen, thin-aired Mars. Impacts of large comets and asteroids are rare enough that mass extinctions are considered extraordinary events. And conditions on the earth have remained hospitable to life for billions of years.

Computer models are beginning to clarify the convoluted circumstances that have led to the earth's happy denouement. George W. Wetherill of the Carnegie Institution of Washington has developed detailed electronic simulations of the final stages of planetary formation, when "planetary embryos"—objects roughly the size of the moon—come crashing together and the final layout of the planetary system becomes clear. Although he admits that his work lies "on the hairy edge of science," Wetherill found that the formation of earthlike planets seems to be the rule rather than the exception.


NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

***GIANT JUPITER** may have helped keep life peaceful on the earth by clearing the solar system of most of its comets. Other planetary systems may not have been so lucky.*

According to current theory, planetary systems form in flattened disks of gas and dust surrounding infant stars through a bottom-up process. Tiny particles coagulate into ever larger bodies, which, aided by their mutual gravitation, rapidly pull together into full-fledged planets. After a few hundred runs of his simulation on a VAX workstation, Wetherill found that most of the time a planet of approximately one earth mass formed between 0.8 and 1.3 times the earth's distance from the sun. (That distance, equal to 149.6 million kilometers, is often referred to as an astronomical unit, or simply AU.)

Not all such planets would necessarily be habitable, of course. Wetherill discovered that conditions on earthlike worlds may be surprisingly dependent on the existence of massive, Jupiter-like planets in the outer solar system. Astronomers think that shortly after the formation of the earth, the giant planets (primarily Jupiter) ejected trillions of comets from the inner part of the solar system and flung most of them into interstellar space. As far as life is concerned, comets can be serious troublemakers; the impact of a large comet is considered one of the likely causes of mass extinctions, such as the one that marked the demise of the dinosaurs.

Building Jupiter turns out to be a tricky problem. A planet can grow to the size of Jupiter only after it acquires enough mass to feed directly off the gas in the nebula surrounding a young star, but the nebula may often dissipate before the protoplanet reaches that critical point. "You might not get Jupiters in many planetary systems," Wetherill observes.

According to Wetherill's model, if Jupiter had failed to form in the solar system, many more comets would have remained in orbits that could eventually bring them into collision with the earth. In that case, impacts would occur about 1,000 times as often as they do in reality. Major extinctions then might happen every 100,000 years or so, causing evolution to take on a very different tack—assuming life managed to gain a toehold at all. "It would make things difficult," Wetherill notes dryly.

Wetherill's equations do not yet prove that Jupiter-class planets are rare. "I wouldn't give up too easily," he urges, noting that in the one planetary system that scientists can study, a Jupiter did manage to form. "The only way to solve the puzzle is to look at other solar systems—something I hope we'll be able to do soon," he says.

Even with comets safely out of the way, Wetherill's calculations do not give any information about whether surface conditions on his earth-size worlds would be habitable. James F. Kasting of Pennsylvania State University, Daniel P.

Whitmire of the University of Southwestern Louisiana and Ray T. Reynolds of the NASA Ames Research Center are investigating the issue by means of computer models designed to simulate climate under various conditions.

Their basic goal is to define the "habitable zone" around a star, that is, the region where a planet would have temperatures that could sustain liquid water and, in principle, life as we know it. "If you can combine what I do with what Kasting does, then we really get somewhere," Wetherill says.

In a recent paper in *Icarus,* Kasting and his collaborators derived the width of the habitable zone around the sun and other similar stars. The inner edge of the habitable zone is defined primarily by the increased rate at which water escapes into the stratosphere, where radiation from the star splits it into oxygen and hydrogen. The researchers found that planets less than 0.95 AU from the sun would have lost their entire water supply over the 4.6-billion-year age of the solar system. Such worlds would be unsuitable for water-dependent forms of life.

At the outer edge of the habitable zone, the main problem is one of keeping warm. A mild greenhouse effect helps the earth to maintain its comfortable temperature. Farther from the sun, a more intense greenhouse effect is needed. Kasting's calculations show that on a planet more than about 1.37 AU from the sun, carbon dioxide begins to freeze in the upper parts of the atmosphere, reflecting more radiation back into space and lowering the temperature still further. This feedback would place the planet in a deep freeze.

Kasting and his co-authors emphasize that their calculations probably underestimate the breadth of the habitable zone. They point to the example of Mars, which lies 1.52 AU from the sun. Ancient channels on the red planet's surface may indicate that nearly four billion years ago the surface was warm enough to permit large bodies of liquid water. That is all the more remarkable because, according to theories of stellar evolution, the sun was roughly 25 percent dimmer then than now. "Early Martian climate is an unsolved problem," Kasting says.

Likewise, the early earth received only a paltry supply of sunlight, yet sedimentary rocks testify to the widespread presence of liquid water at least 3.8 billion years ago. One possible explanation, embraced by I.-Juliana Sackmann of the California Institute of Technology and several others, is that the early sun was more massive, and hence brighter, than conventional theory predicts. Kasting favors a less radical but "still speculative" notion that the atmosphere of the young earth contained traces of extremely effective greenhouse gases such as ammonia and methane.

Somehow the earth, Jupiter and the sun managed to develop in precisely the right way so that terrestrial conditions always remained suitable for water-based life. Theoretical models represent the first step in determining whether the earth is just a lucky fluke. Perhaps the shape of the solar system is the most logical consequence of the way planetary systems form. "It could be a natural, self-regulated machine," Wetherill muses. In that case, the numbers spit out by his electronic simulations may correspond to a multitude of real, habitable worlds. —*Corey S. Powell*

What If They Don't Have Radios?

Are mathematical theorems and theories of physics universal truths, likely to be discovered by any beings given to pondering the nature of things? Or are they inventions, as much products of our idiosyncratic heritage and needs as eyeglasses or toasters?

This old conundrum could be put to a test of sorts by the National Aeronautics and Space Administration's ambitious new search for intelligent life elsewhere in the universe. Called the High Resolution Microwave Survey (the old name, the Search for Extraterrestrial Intelligence, or SETI, was scrapped because it was thought to evoke science fiction rather than science), it involves scanning the heavens for alien radio signals.

So far NASA has dedicated two telescopes to the effort. The 305-meter fixed dish at Arecibo, Puerto Rico, is tuning in to a select group of stars within 100 light-years of the earth, and a 34-meter movable dish at Goldstone, Calif., is sweeping broad swaths of the sky. NASA hopes to continue the effort for at least 10 years, for a total cost of $100 million.

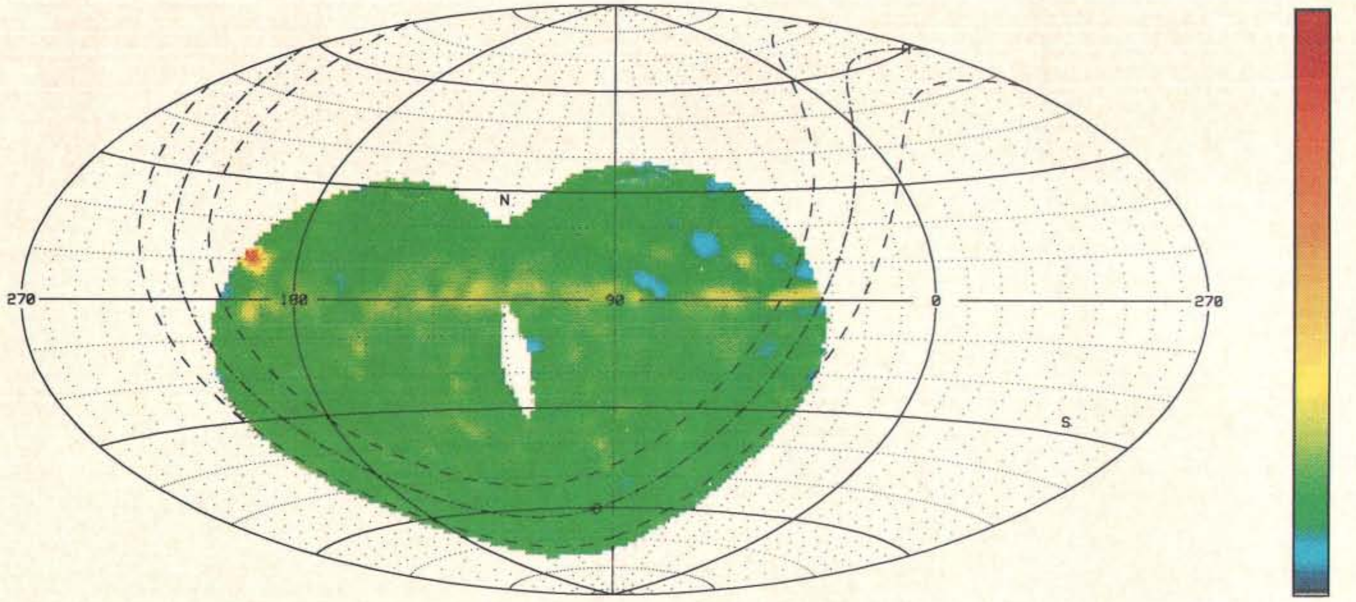
Why would workers expect either instrument to detect signs of intelligent life? Because, explains Frank D. Drake, a physicist at the University of California at Santa Cruz and a veteran SETI researcher, intelligent extraterrestrial beings would have "basically the same" systems of mathematics and physics that we have. "Many human societies developed science independently through a combination of curiosity and trying to create a better life," he notes, "and I think those same motivations would exist in other creatures."

Inevitably, he argues, alien scientists would discover gravity, electromagnetism and other fundamental physical phenomena. It follows that they would develop technologies such as radio communications. Drake also thinks intelligent aliens are likely to discover such esoteric concepts as the theory of general relativity, quantum-field theory and even superstrings.

This view is "infinitely parochial," argues Nicholas Rescher, a philosopher at the University of Pittsburgh. "It's like saying they would have the same legal or political system that we do." Rescher contends that our science, mathematics and technology are unique outgrowths of our physiology, cognitive makeup and environment. Indeed, the whole SETI enterprise is "a waste of time, money and energy," Rescher says. "It's perfectly possible that there are other civilizations, and it's perfectly possible that they communicate in some way. But that they communicate in the same basic way we do is about as likely as it would be that they communicate in English."

An intermediate point of view is offered by John D. Barrow, an astronomer at the University of Sussex in England. Barrow, author of a new book, *Pi in the Sky,* that explores the issue of whether mathematics is discovered or invented, believes aliens may well share some basic ideas underlying mathematics and physics, such as the concepts of counting or of cause and effect. "There are certain aspects of the world that press themselves on us," he says.

But as science becomes more removed from everyday reality, Barrow notes, its development may become more serendipitous. The theory of relativity, for example, became accepted only after observations of a solar eclipse confirmed Einstein's prediction about the bending of light. Those observations were possible because the sun and the moon, as seen from the earth, are almost exactly the same size. Actually, Barrow is more concerned about the ethics of little green men than about their science. If we meet aliens, will they have the equivalent of the Golden Rule: Do unto others as you would have them do unto you? —*John Horgan*



MICROWAVE MAP derived from the M.I.T./Princeton balloon experiment matches observations by the Cosmic Background Explorer *satellite. Coolest regions are blue, and warmest are red. The red spot at the left is Jupiter.*

COBE Corroborated

Balloon observations support satellite data

When a team of investigators announced last April that the *Cosmic Background Explorer* (*COBE*) satellite had discovered minute fluctuations in a faint glow of microwaves left over from the big bang, cosmologists were understandably overjoyed. Lacking evidence of inhomogeneity, they would have been hard-pressed to explain how the early universe evolved into its current, rather lumpy condition. Yet their exultation was tinged with anxiety. The signals detected by the *COBE* team were barely discernible through the ambient noise. What if they were illusory?

Now those fears have been greatly allayed by data from a balloon-borne instrument that soared aloft from New Mexico for 12 hours in 1989. In December participants in the M.I.T./Princeton microwave background experiment finally announced during a workshop at the University of California at Berkeley that they had corroborated *COBE*'s results.

Unlike *COBE*, which surveys the entire sky, the balloon experiment mapped only a third of the sky. But the map presented by Stephan S. Meyer of the Massachusetts Institute of Technology, Lyman A. Page and Kenneth M. Ganga of Princeton University and Edward S. Cheng of the NASA Goddard Space Flight Center shows fluctuations whose amplitude and overall pattern match those of *COBE*. "Smoot seems to be very happy," Meyer said, referring to George F. Smoot, a leader of the *COBE* team.

The balloon team turned up hints of the cosmic fluctuations by 1991. But they still had to rule out the possibility that the signals had come from noncosmic sources. The workers were able to pinpoint and thus eliminate radiation from the Milky Way by comparing their map with one made by the *Infrared Astronomical Satellite* (*IRAS*).

Systematic errors in the instruments could also have created spurious features, but the agreement between the data from the balloon and from *COBE* makes that possibility unlikely, according to Meyer. "Systematic errors of the same size in two different experiments would be very rare," says Meyer, who is also a member of the *COBE* team.

Even so, the balloon map, like the *COBE* one, is highly probabilistic in nature. In other words, investigators cannot assert with certainty that any particular feature in either map actually exists or is a statistical artifact. *COBE* should have gathered enough data to rectify that situation within another year or so, Meyer says.

One of the drawbacks of the *COBE* and M.I.T./Princeton maps is that their resolution is very broad. Indeed, the cosmic features they have detected are huge, larger than even the largest voids and superclusters of galaxies detected so far by optical telescopes. For that reason, theorists have been eagerly awaiting results from two other probes of the microwave background: the *Advanced Cosmic Microwave Explorer* (*ACME*), which involves ground-based measurements made at the South Pole, and the *Millimeter Anisotropy Experiment* (*MAX*), which consists of balloon-based observations.

ACME and *MAX* scan swaths of sky about 10 times smaller than those examined by *COBE* and the M.I.T./Princeton groups. The finer-scale observations should be "more directly relevant to structure formation," says Philip M. Lubin of the University of California at Santa Barbara, a member of the *ACME* and *MAX* (and *COBE*) teams.

Both groups have glimpsed small-scale fluctuations in the microwave background, according to Lubin. He emphasizes that more observations are needed to eliminate the possibility that radiation from the Milky Way or other galaxies caused the fluctuations. "It may be cosmological, or it may be galactic, so we won't bring our fist down hard on the table yet," he says.

Of course, theorists cannot resist interpreting these preliminary results. So far their glosses have favored two related hypotheses that have been rather battered lately: inflation, which holds that the early universe passed through a prodigious growth spurt, and cold dark matter, which posits that the universe is composed for the most part of slow-moving, difficult-to-detect matter. "If Lubin's fluctuations are the real thing," notes Joseph I. Silk, a theorist at Berkeley, "then inflation and cold dark matter look very nice." —*John Horgan*

Genes and Crime

A U.S. plan to reduce violence rekindles an old controversy

No one disputes that such environmental factors as poverty, unemployment and drugs contribute to the high rates of violent crime plaguing the U.S. Agreement dissolves, however, when the possibility is put forward that some people are born with an innate predisposition toward violent crime. This issue, which has long lurked at the fringes of respectable scientific discourse, has been thrust into prominence during the past year by a planned federal antiviolence initiative.

The initiative was conceived more than a year ago by Louis W. Sullivan, then secretary of health and human services. As a black physician, Sullivan explicitly intended the initiative to help blacks, who are disproportionately affected by violent crimes. The black homicide rate is five times higher than is the rate for whites, and homicide is the major cause of death of black males between the ages of 15 and 24. Blacks are also six times more likely to be arrested for a violent crime than are whites.

The five-year, $400-million program planned by Sullivan would integrate and boost federal funding for violence research, now at about $50 million a year. Most of the research, Sullivan has repeatedly emphasized, would be "psychosocial," examining child abuse, drug addiction and other potential causes of crime. The program would also evaluate preventive measures such as counseling and gun control. Only about 5 percent of the initiative's budget would fund "biological" research, including studies of hormones and neurotransmitters linked to aggressive behavior in animals and humans.

Yet controversy over this aspect of the initiative was triggered last year by Frederick K. Goodwin, then director of the Alcohol, Drug Abuse and Mental Health Administration. Goodwin, who now heads the National Institute of Mental Health, cited research on monkey violence and sexuality and commented that "maybe it isn't just the careless use of the word when people call certain areas of certain cities 'jungles.' "

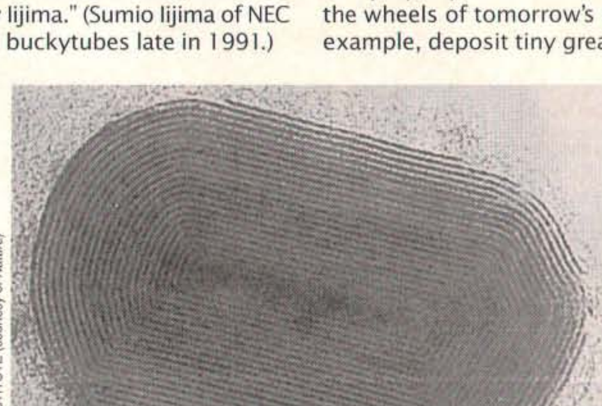
Civil rights leaders and others were still fuming last summer when announcements were mailed out for a conference titled "Genetic Factors in Crime:

Faux Fullerenes

Like a newly learned word that seems to jump from every book, molecular cages have become ubiquitous since the existence of buckminsterfullerene's icosahedral carbon cage was confirmed two years ago. First came larger carbon cages, called giant fullerenes; nested cages, known as Russian dolls; and ultrathin fibers, called buckytubes. Next were the metallofullerenes—hybrids that encase metal atoms or incorporate them in the carbon lattice itself. Now the synthesis of a carbonless envelope has been announced: a nested cage of tungsten disulfide [*see illustration below*].

The faux fullerenes first appeared in July 1991 at Israel's Weizmann Institute of Science, where Reshef Tenne, Lev Margulis, Menachem Genut and Gary Hodes were preparing tungsten disulfide for use in high-performance solar cells. The workers did not immediately grasp the importance of the nested balls of the semiconductor material. "We saw the Russian dolls in July 1991," Tenne says, "but we did not make the connection until later, when we looked at the pictures made by Iijima." (Sumio Iijima of NEC Corporation described nested buckytubes late in 1991.)

As a result of the delay, the Weizmann researchers can state categorically that the mock buckystructures are stable for at least a year. But easy though they may be to keep, no one has yet produced them in bulk. Like their carbon archetypes, phony fullerenes form only at high temperatures. In such a regime, a vapor of tungsten disulfide condenses into a two-dimensional sheet, as do the carbon precursors of fullerenes. Some hexagonal cells then convert to pentagons, causing the sheet to curve in on itself and close.



WEIZMANN INSTITUTE (courtesy of *Nature*)

UNCARBONATED FULLEROID consists of nested cages of the semiconductor tungsten disulfide.

What tricks might these motes perform if they could be made by the gram? "I guess they will show photoconductive and quantum effects," Tenne says. The smallest cages of tungsten disulfide are believed to have an electronic band gap well below the 1.6 electron volts of the bulk material. "As the number of layers rises," Tenne notes, "the gap should approach that value." Materials scientists can therefore hope to control the growth of the structures so as to "tune" the band gap for their electronic properties. For example, one might tune the Russian dolls for optimal absorption of sunlight, producing better solar cells. Even more exciting is the prospect of tuning tungsten disulfide so that it emits visible light. The bulk form of this material cannot serve this function, because it is, like silicon, an indirect-gap semiconductor, in which electrons and positive charges, or holes, do not normally recombine to form light.

Other possibilities also beckon. Tungsten disulfide is used as a lubricant in some aerospace applications. If it retains this property in its fulleroid form, it may serve to grease the wheels of tomorrow's nanomachines. One might, for example, deposit tiny greaseballs in a microscopic bushing or inside a minuscule ball-and-socket joint. Mock buckytubes might also be intercalated with lithium to form microscopic, rechargeable batteries.

The range of properties of fakeyballs looms even larger because other substances can also condense into sheetlike precursors. Each substance might father an entire family of shapes and sizes. "Oh, there are so many two-dimensional materials," Tenne exults. "We are trying molybdenum disulfide. Then we will go to other compounds." —*Philip E. Ross*

Findings, Uses and Implications," which was to be held at the University of Maryland in October. The conference brochure noted "the apparent failure of environmental approaches to crime" and suggested that genetic research might lead to methods for identifying and pharmaceutically treating potential criminals at an early age.

David T. Wasserman, a legal scholar at the University of Maryland and organizer of the meeting, insisted it was intended to critique rather than promote this view, but critics were not mollified. Peter Breggin, a Bethesda-based psychiatrist, linked the Maryland conference to Goodwin's remarks and to the violence initiative. The U.S., he proclaimed, was planning a large-scale program to screen black children and treat them with drugs. "U.S. government wants to sedate black youth," announced a black-interest magazine in Washington, D.C.

A committee of the National Institutes of Health had already approved funds for the conference. But in response to the criticism, NIH director Bernadine P. Healy withheld the funds, and the meeting was indefinitely postponed. But then in November the National Academy of Sciences issued a 464-page report, "Understanding and Preventing Violence," calling for more research of the kind that the Maryland conference would have examined, including searches for biochemical markers and drug treatments for violent and antisocial behavior.

Given the inexorable advance and acceptance of genetics research, the debate is likely to intensify, according to Diane B. Paul, a political scientist at the University of Massachusetts at Boston. "We are more and more focused on genetics," explains Paul, who is skeptical of research linking genes to behavioral disorders. "When [former head of the Human Genome Project] James D. Watson says, 'We used to think our fate was in the stars, and now we know it's in our genes,' he's giving expression to a social current."

Of course, claims of links between heredity and crime have a long and sordid history. Some Victorian-era scientists contended that criminals were more likely to have small, shifty eyes, eyebrows that met in the middle and other traits. Through the 1930s, many U.S. states—with the sanction of the Supreme Court—sterilized convicts in order to reduce crime among future generations. More recently, some prominent scientists, notably Richard J. Herrstein, a psychologist at Harvard University, have suggested that blacks may be intrinsically more prone toward criminal behavior than whites are.

Although most scientists reject these views, many have been convinced by studies of adoptees and other populations that heredity influences virtually all aspects of human behavior—from intelligence to sexual orientation. Buoyed by the successful identification of genes responsible for cystic fibrosis, Duchenne's muscular dystrophy and other diseases, researchers are now looking for genes associated with such disorders as alcoholism, schizophrenia and manic depression.

The NAS report acknowledges the paucity of substantive evidence for a genetic propensity for crime per se. The most frequently cited study was done a decade ago by Sarnoff A. Mednick of the University of Southern California. Comparing the criminal records of some 14,000 adopted Danish males with the records of their biological and adopted fathers, Mednick found evidence of heritability—but only for property crimes, for example, burglary, and not for violent crimes. On the other hand, studies involving adopted children have yielded tentative evidence of a genetic influence underlying traits sometimes associated with crime, among them aggressiveness, impulsiveness and susceptibility to addiction.

No one has claimed that there may be a "crime gene" that could serve as a marker and perhaps even be manipulated for therapeutic purposes. Crime is, after all, an extremely heterogeneous—and culturally defined—phenomenon. But some scientists have proposed that it might be possible to find physiological markers for certain crime-related attributes. The most popular current candidate for a marker is the neurotransmitter serotonin. Studies of animals and humans indicate that as levels of serotonin decrease, the propensity for aggression and violence increases.

To be sure, not all investigators of violence and criminality accept that heredity plays any significant role. "Those who study genetic components generally fail to look at the social and psychological variables," says Joan McCord, a sociologist at Temple University. McCord analyzed data from a long-term study of 34 pairs of brothers born in the Boston area between 1926 and 1933. Comparing the criminal histories of brothers with each other and with the histories of subjects having similar backgrounds, McCord found no significant evidence for a genetic contribution to criminality.

Most of the subjects in the Boston study were white. McCord opposes studies that interpret differences in terms of race, arguing that race is a social and not a biological category. But avoiding race in studies of violence and crime is "playing into the hands of the right wing," says Troy Duster, a sociologist at the University of California at Berkeley. If studies properly account for racism and related factors, allegations of a black propensity for criminality "will fade away into nothing," he says.

Indeed, a recent study by C. Robert Cloninger of Washington University supports this view. Cloninger examined the prevalence of personality factors that have been shown to be heritable and associated with criminality later in life—including impulsivity and aggressiveness—in more than 1,000 adults of various races. He found essentially the same proportion of crime-linked traits in both the white and black populations. The higher rates of criminality observed among blacks, Cloninger concludes, "must be the result of socioeconomic factors or other environmental variables."

Ronald W. Walters, a political scientist at Howard University who led the fight against the canceled University of Maryland conference and is a founding member of the National Committee to Stop the Violence Initiative, opposes all research on the biological causes of crime. "There are some things you're better off not to know if you're going to live together," he says.

"I very strongly object to anybody who says knowledge is dangerous," responds Kenneth K. Kidd, a geneticist at Yale University. "Notice I said knowledge and not theories spouted off." Kidd, who has been involved in the search (fruitless so far) for specific genetic markers for mental illnesses, says he has no doubt that genes play some role in criminal behavior.

On the other hand, Kidd questions the value of research into genetic factors for crime since it is unlikely that researchers will ever isolate genes associated with such a complex phenomenon. "If one can come up with a good definition of a type associated with extreme violence, I'd say, fine, let's try to understand that."

The debate has spurred the American Association for the Advancement of Science to schedule a session called "Controversy over Crime and Heredity: An Exploration" for its annual meeting, to be held in Boston in February. Robert F. Murray, a geneticist at Howard who is moderating the AAAS session, hopes it will quell some of the "hysteria" surrounding the violence initiative, which he supports. Even so, Murray admits to misgivings: "My concern is that the research will be used not for people's benefit but to denigrate or stigmatize them." *—John Horgan*

The Artist, the Physicist and the Waterfall

Roger Penrose, now a professor at the University of Oxford, was a 23-year-old graduate student when he encountered the geometric art of Maurits C. Escher at a mathematics conference in Amsterdam in 1954. Since then, the British mathematician and physicist seems to have shared a mysterious, space-and-time-transcending bond with the late Dutch artist.

Like many mathematicians, Penrose was fascinated by Escher's playful exploration of such concepts as symmetry and infinite regress—and his manipulation of perspective and geometry to construct "impossible" objects, which violate the rules of three-dimensional reality. Escher's drawings inspired Penrose to doodle an impossible object of his own, a "tribar" made of three conjoined beams. The tribar appears straightforward at first, but as one traces its beams one realizes that they—or is it space itself?—must be twisted.

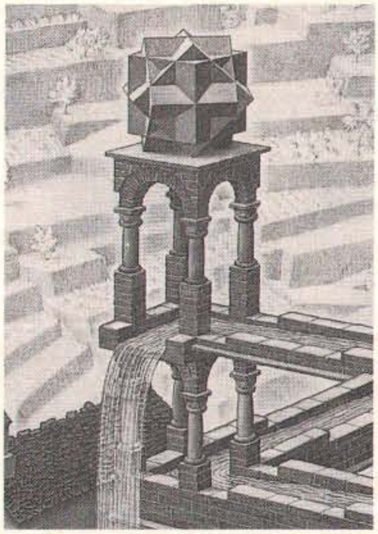
Penrose showed the tribar to his father, Lionel, a prominent geneticist from whom Roger inherited his love of puzzles. Lionel responded by sketching an impossible staircase, one that seems to ascend but somehow keeps circling back on itself. Together father and son wrote a paper describing the triangle and staircase and sent it to Escher. The paper, published in the *British Journal of Psychology* in 1958, spurred Escher in turn to create two of his most famous lithographs: *Ascending and Descending*, which depicts monks tramping up and down a Sisyphean staircase, and *Waterfall*, which transforms Roger's tribar into a perpetually flowing circuit of water.

The story resumes three decades later, in May 1991, when Penrose attended a meeting in Copenhagen on quantum physics. There he heard the physicist Asher Peres of Technion University in Israel lecture on hidden-variable theories. These theories attempt to explain quantum effects such as nonlocality—in which particles emitted by a common source influence one another across vast distances—in classical terms, by invoking undetectable forces or properties. Peres proposed that one can unambiguously rule out a broad class of hidden-variable theories by measuring the spin of a particle with respect to 33 directions, defined by coordinates in three dimensions.

Penrose, who often tries to envision concepts in geometric terms, asked Peres if his coordinates corresponded to any interesting polyhedrons. "He just looked at me blankly," Penrose recalls. "So I decided I'd draw some pictures and see if they made any sense." Sure enough, as Penrose plotted Peres's coordinates, a complex polyhedron emerged on the page. It consisted of three interpenetrating cubes, each rotated 90 degrees with respect to the others. "I looked at it," Penrose says, "and I thought, 'Gosh, I've seen that somewhere before.' " Suddenly he remembered: Escher had set just such a polyhedron atop the left-hand tower of his waterfall structure. Penrose has written up his "curious" finding for a volume of papers to be published in memory of the great quantum theorist John Bell. Unfortunately, Penrose cannot send the paper to Escher, because the artist died 21 years ago.

Penrose did meet Escher once, in 1962. "I happened to be driving in Holland," he recalls, "so I phoned him up, and he invited me over for tea." Penrose presented Escher with a puzzle: a set of identical polygons that, if fitted together properly, could generate an infinite plane. Escher later solved the puzzle—the key was flipping over some polygons to turn them into mirror-symmetric counterparts—and in 1971, just before he died, he drew a picture based on the puzzle.

In one respect, the encounter was a bit disappointing. "I thought his house might have a staircase going out the window or something," Penrose remarks. "But everything was very neat and organized." —*John Horgan*


M.C. ESCHER *Cordon Art*

***QUANTUM POLYHEDRON adorns a tower in M. C. Escher's* Waterfall.**

Electronic Envelopes?

The uncertainty of keeping e-mail private

Recent legislative efforts to mandate remote wiretapping attachments for every telephone system and computer network in the U.S. may have been the best thing that ever happened for encryption software. "We have mostly the FBI to thank," says John Gilmore of Cygnus Support in Palo Alto, Calif. Gilmore is an entrepreneur, hacker and electronic civil libertarian who helped to found the Electronic Frontier Foundation (EFF). He is now watching closely the development of two competing techniques for keeping electronic mail private.

As matters now stand, computers transmit messages from one user to another in plain text. If a geneticist in Boston sends e-mail to a molecular biologist in San Diego, any of the half a dozen or so intermediary machines that forward the letter could siphon off a copy—and so could any of the dozens of workstations that might be attached to the local-area network at the sender's or recipient's university or company.

The Electronic Privacy Act of 1986 prohibits snooping by public e-mail carriers or law-enforcement officials, except by court order. Nevertheless, many people are becoming uncomfortable with the electronic equivalent of mailing all their correspondence on postcards and relying on people to refrain from reading it. They are turning to public-key encryption, which allows anyone to encode a message but only the recipient to decode it. Each user has a public key, which is made widely available, and a closely guarded secret key. Messages encrypted with one key can be decrypted only with the other, thus also making it possible to "sign" messages by encrypting them with the private key [see "Achieving Electronic Privacy," by David Chaum; SCIENTIFIC AMERICAN, August 1992].

Two programs—and two almost diametrically opposed viewpoints embodied in them—are competing for acceptance. Privacy Enhanced Mail (PEM) is the long-awaited culmination of years of international standard setting by computer scientists. Pretty Good Privacy (PGP) is a possibly illegal work of "guerrilla freeware" originally written by software consultant Philip Zimmermann.

The philosophies of PEM and PGP differ most visibly with respect to key management, the crucial task of ensuring that the public keys that encode messages actually belong to the intended recipient rather than a malevolent third

party. PEM relies on a rigid hierarchy of trusted companies, universities and other institutions to certify public keys, which are then stored on a "key server" accessible over the Internet. To send private mail, one asks the key server for the public key of the addressee, which has been signed by the appropriate certification authorities. PGP, in contrast, operates on what Zimmermann calls "a web of trust": people who wish to correspond privately can exchange keys directly or through trusted intermediaries. The intermediaries sign the keys that they pass on, thus certifying their authenticity.

PGP's decentralized approach has gained a wide following since its initial release in June 1991, according to Hugh E. Miller of Loyola University in Chicago, who maintains an electronic mailing list for discussion among PGP users. His personal "keyring" file contains public keys for about 100 correspondents, and others have keyrings containing far more. As of the end of 1992, meanwhile, a final version of PEM had not been officially released. Gilmore, who subscribes to the electronic mailing list for PEM developers, says he has seen "only five or 10" messages actually encrypted using the software.

Although PGP's purchase price is right—it is freely available over the Internet and on electronic bulletin boards throughout the world—it does carry two liabilities that could frighten away potential users. First, U.S. law defines cryptographic hardware and software as "munitions." So anyone who is caught making a copy of the program could run afoul of export-control laws. Miller calls this situation "absurd," citing the availability of high-quality cryptographic software on the streets of Moscow.

Worse yet, RSA Data Security in Redwood City, Calif., holds rights to a U.S. patent on the public-key encryption algorithm, and D. James Bidzos, the company's president, asserts that anyone using or distributing PGP could be sued for infringement. The company has licensed public-key software to corporations and sells its own encrypted-mail package (the algorithm was developed with federal support, and so the government has a royalty-free license). When Bidzos's attorneys warned Zimmermann that he faced a suit for developing PGP, he gave up further work on the program.

Instead PGP's ongoing improvements are in the hands of an international team of software developers who take advice from Zimmermann by e-mail. The U.S. is the only nation that permits the patenting of mathematical algorithms, and so programmers in the Netherlands or New Zealand apparently have little to fear.

U.S. residents who import the program could still face legal action, although repeated warnings broadcast in cryptography discussion groups on computer networks have yet to be superseded by legal filings. Meanwhile, Gilmore says, the only substantive effect of the patent threat is that development and use of cryptographic tools have been driven out of the U.S. into less restrictive countries. —*Paul Wallich*

Making Waves

As the principals in the cold war slowly dismantle their land-based missiles, submarine-based ballistic missiles are assuming greater importance as ultimate deterrents. Because submarines are undetectable in the vast ocean basins, the theory goes, the fear of inevitable retaliation would discourage an attack on a state possessing them.

The effectiveness of that deterrent may now be in doubt. Russian scientists who specialize in remote-sensing research have been making waves by claiming to have demonstrated a way of detecting submerged ballistic-missile submarines, using microwave reflections from the sea surface. The Russians say they have described their findings to their U.S. counterparts in the hope of avoiding a breakthrough by either side alone that might jeopardize the "build down" of strategic weapons.

The Russian claims are based on the work of Valentin S. Etkin, head of the applied space physics department at the Space Research Institute in Moscow, where the Russian work is concentrated. Etkin has pursued a line of thinking that others have entertained before him: internal waves in the ocean at the boundaries between layers of different density seem likely to cause subtle disturbances at the surface.

Researchers who have looked for such effects as a way of detecting internal waves caused by submarines have typically used visible light or infrared radiation. Etkin, in contrast, maintains that he can detect the changes by looking at reflected microwaves. "This is extremely important and promising," says Vyacheslav M. Balebanov, deputy director of the Space Research Institute. "In principle, it's not so difficult to see a submarine at less than 100 meters, and we have got positive results."

Etkin will not discuss what his results mean for submarine detection, but he has reportedly described spotting several submarines at substantial depths to U.S. military personnel. As a result, the Senate Armed Services Committee put pressure on the Pentagon to investigate Etkin's work. "Never before has a foreign state proposed to demonstrate to the United States that it can detect our submarines at sea," the committee notes in a report on the START treaties.

The U.S. Navy and the Central Intelligence Agency, anxious not to appear complacent, decided to sponsor a joint research project with Etkin. Both have invested huge sums in a fruitless effort to find ways of detecting submarines, according to Clarence A. Robinson, the editor of *Signal*, a journal of military communications.

Researchers from the Applied Physics Laboratory at Johns Hopkins University and from the Space Research Institute conducted several days of observations last summer in an area of the Atlantic Ocean 60 miles off New York City, called the Long Island Bight. The unclassified effort, sardonically designated CHERI (Critical Hydrodynamic and Electrodynamic Research Issues), sought to observe the effects on the sea surface not only of internal ocean waves but also of waves and convection cells in the atmosphere, Etkin says.

A Russian airplane, research ship and satellite took part, as did two U.S. aircraft. From these platforms, different combinations of radar and other devices were directed at the ocean. A continuation of the experiment is planned for this year in the Pacific Ocean off Russia's Kamchatka peninsula.

Michael Kobrick, a National Aeronautics and Space Adminstration scientist who participated in the New York project, says the radar clearly detected waves beneath the surface, although no submarines were used in the experiment, according to Kobrick. Even though General Colin L. Powell, chairman of the joint chiefs of staff, told the Senate last year that the U.S. is trying to conduct the collaboration without revealing secrets, Etkin, for one, says he is enthusiastic about continuing the work. —*Tim Beardsley*

PROFILE: NATHAN P. MYHRVOLD

The Physicist as a Young Businessman

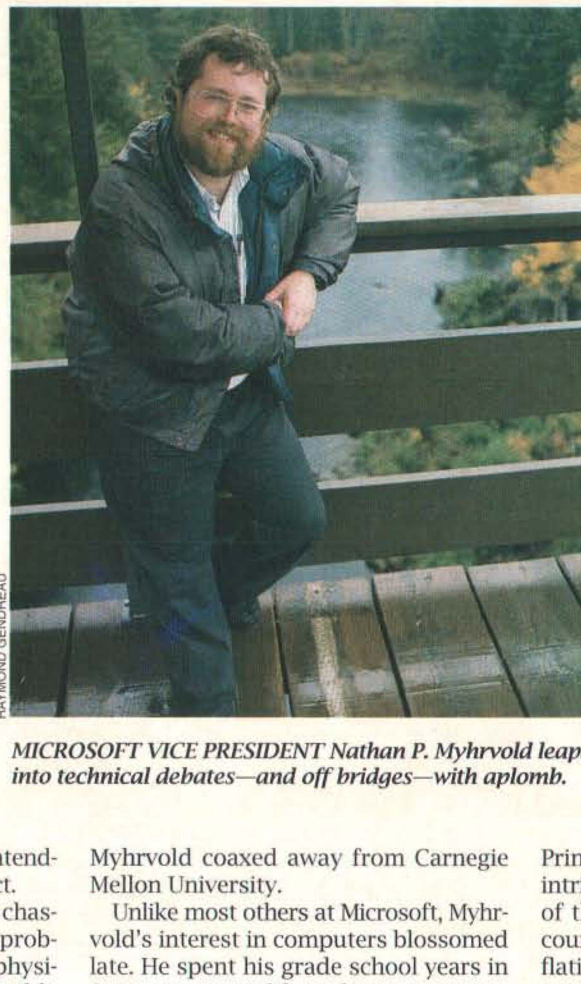
Nathan P. Myhrvold leans back in his chair, arms folded behind his head, legs stretched out. A mop of auburn curls tumbles around wire-frame glasses, running into a thick beard. His shirt, hastily tucked into gray chinos, threatens to come undone around the midriff. The ambience is casual; he could be holding after-class office hours at a university or rapping with a partner in a start-up.

Instead this is Myhrvold's weekly tête-à-tête strategy session with Microsoft chairman William H. Gates. And Myhrvold, the company's 33-year-old vice president for advanced technology and business development, has brought a serious agenda: new markets, possible acquisitions and the plans for the company's evolving research laboratory. The young, erstwhile physicist stands just outside the spotlight but close to the helm. "Other than myself, Nathan has more impact on our long-term strategy than anyone else," Gates says.

That responsibility seems to rest comfortably on the shoulders of someone who once spent his time trying to unravel the origins of the universe. Myhrvold joined Microsoft in Redmond, Wash., less than seven years ago, with one start-up company and a brief stint with physicist Stephen W. Hawking already on his résumé. He now carries direct responsibility for creating both Microsoft's research laboratory and a parallel "advanced development" group, intended to help propel ideas into product.

Myhrvold has spent his career chasing "very hard but not insoluble problems," driven by the bravura of a physicist and tempered by an irrepressible sense of humor. Last autumn when the local United Way chapter asked Microsoft executives to volunteer for activities that could be "auctioned off" in a fund-raising event, many offered to go see a movie or a ball game with the successful bidder. Not Myhrvold. "I'm going to jump off a bridge," he brightly told a visitor. He had never tried "bungee jumping," in which the intrepid jumpers are snatched from death's door by elastic cords. But for United Way, he happily leapt off a bridge in Vancouver.

Those who have joined the research group have found Myhrvold a surprisingly accessible boss. "What other company offers this kind of perk, where your boss will spend an entire day slaving over a hot barbecue, cooking your dinner?" demands Richard F. Rashid, whom Myhrvold coaxed away from Carnegie Mellon University.



RAYMOND GENDREAU

MICROSOFT VICE PRESIDENT Nathan P. Myhrvold leaps into technical debates—and off bridges—with aplomb.

Unlike most others at Microsoft, Myhrvold's interest in computers blossomed late. He spent his grade school years in Santa Monica, Calif., soaking up science, particularly ideas in mathematics and biology. By age 14, the only course separating him from a high school diploma was driver's education. He filled the next two years by taking classes at Santa Monica City College and wound up graduating in 1979 from the University of California at Los Angeles with a bachelor's degree in mathematics and a master's in geophysics and space sciences. A year or so later he added a second master's degree to his collection, this time in mathematical economics from Princeton University.

When it came time to choose a topic for his doctoral thesis, Myhrvold cast about for one that demanded even more intellectual acrobatics. "I thought general relativity was pretty cool, and quantum-field theory, curved space time and quantum gravity would be cooler yet," he recalls. So he wrapped them up together and began looking for a way to explain gravity in the context of quantum mechanics. "It's a nearly impossible problem," he says cheerfully—but also a profound one. "I used to be able to get all worked up about it and say, 'What other possible thing should someone study?' "

Tackling cosmology is about as "blue sky" as science gets. Even scarcer than the physical evidence for theories are full-time jobs for the scientists who pursue them. The very nature of the challenges fuels a streak of intellectual machismo. "It takes this funny sort of ego, you know, playing chicken with Mother Nature," Myhrvold says.

"By the time you get to graduate school you've gone through *n* people saying, 'Look, you really should consider something more practical,' " Myhrvold notes. "But if they follow it up with something like, 'Only the most brilliant people get by,' are you going to say, 'Yeah, I'm going to wimp out on this one'?"

Myhrvold's thesis, "Vistas in Curved Space Quantum Field Theory," which he completed at Princeton, offers a still unproved but intriguing proposal for tidying up some of the loose ends in the "inflation" account of the origins of the universe. Inflation theorists suggest that the universe began as a highly curved core of matter. At some point, that core exploded, exponentially expanding the size of the universe and spewing forth stars and planets. What started the expansion and why it stopped remain subjects of debate. "My mechanism turned it off," he proposes.

Yet even as he was finishing his thesis, Myhrvold's attention began stray-

ing to computers. A few scientists, including Stephen Wolfram, then at the Institute for Advanced Studies in Princeton, had begun writing computer programs that could manipulate equations with thousands of terms. Myhrvold and a few fellow students, in contrast, began tinkering with a program they envisioned could be a kind of word processor for mathematicians—software that would deftly juggle the simpler equations that scientists use routinely.

Around the same time, Myhrvold won a postdoctoral position with Hawking at the University of Cambridge. He spent about a year in England working on such problems as finding a wave function for the entire universe. When summer arrived, Myhrvold took a leave from Cambridge to continue to work on software.

There was, however, a hitch. No operating system would support the kind of symbol manipulation Myhrvold and two fellow Princeton physicists wanted to do on an IBM-compatible personal computer. So the team, joined by Myhrvold's younger brother and another friend, decided to "spend just a little bit of time whipping up something that would be an operating environment, a little 'windows' system," Myhrvold says.

By the end of the summer, their work had tweaked the interest of a few venture capitalists. Myhrvold gave up his position with Hawking; one of his colleagues, another freshly minted physicist, never showed up for his postdoctoral position at the Fermi National Accelerator Laboratory. Working from the attic of a house in Berkeley, Calif., the group formed Dynamical Systems Research, with Myhrvold as president.

Unbeknownst to the fledgling Dynamical Systems, a pack of other software writers were also busily writing similar code. Among them, Microsoft was in the early days of what would become "Windows." IBM was building a similar operating system called "TopView" for its line of advanced personal computers.

When IBM released TopView in 1985, enthusiasm for Dynamical System's work plummeted like a dead bird from a roof. "We drove down to the IBM product center and got a demo," Myhrvold says, sighing. In many ways, TopView offered the same capabilities that Myhrvold's team had hoped to achieve.

So Dynamical Systems chose a risky path: the company "cloned," or mimicked, the functions of TopView, using only a quarter of the memory space. A handful of companies, including Merrill Lynch and a fish-canning plant in Denmark, signed up as customers. But Dynamical Systems continually teetered on the edge of insolvency. Most of its then dozen employees worked largely for stock options. "It's hard not to wonder if you've led them all astray, when the company is down to its last hundred bucks," Myhrvold says.

Hope appeared in the form of Microsoft, which still was trying to coax IBM to buy into its nascent Windows user interface. To sweeten the offer, Microsoft pledged to make Windows compatible with TopView, primarily by adding in elements of Dynamical Systems' software. In the process, Microsoft decided to buy the start-up—code, customers and physicists intact.

Known for his 100-page memos on future technologies, Myhrvold sparks imaginations.

As it turned out, Microsoft never used the software written by Myhrvold and his team. TopView also faded from sight. Myhrvold nonetheless swiftly became engaged in the joint IBM-Microsoft development of the interface for the operating system, OS/2. He also led the graphics team on a parallel project, which became Windows 2.0. By the mid-1980s, the relationship between IBM and Microsoft was already growing fragile. The clashes and disappointments over OS/2 would push the two companies irrevocably apart by 1990. Microsoft continued to promote Windows, and IBM carried OS/2 alone.

Both Microsoft and IBM have reaped their shares of criticism for OS/2. In *Computer Wars,* a book scheduled for publication in February, industry analysts Charles H. Ferguson and Charles R. Morris argue that Gates and Microsoft made technical decisions that crippled OS/2—and knew it at the time.

Myhrvold vigorously denies the charge. "Some people think there was some deliberate bait and switch. It's just absolutely not true," he insists. "I tried very hard. Then came a point when I realized it just couldn't work. So I tried very hard to kill it. The joke at Microsoft is that I offer cradle-to-grave service on systems. Am I proud of that? I couldn't do anything different—other than not work with IBM, and that wasn't feasible for us at the time."

Edward Iacobucci, IBM's design leader for OS/2 and now chairman of Citrix Systems in Coral Springs, Fla., confirms that Myhrvold put his back into OS/2 and the proposed follow-on. Differences in business strategies, not technology, pulled IBM and Microsoft apart, he says.

These days at Microsoft, Myhrvold spends much of his time thinking about the future. His omnivorous curiosity is serving him in good stead as he sets the charter for the research and advanced technology groups, which he expects will top 120 people within two years. He tries to spark the imagination of the cadre of researchers by inviting artificial-life experts and science fiction writers as guest lecturers. "He has a compendium of knowledge," says Edward Jung, a software architect at Microsoft. "And he writes these memos—"

Indeed, memos have become a Myhrvold trademark. Using an ergonomically advantageous Dvorak keyboard, Myhrvold pounds out summaries of industrial trends or visions of the future, many of which number more than 100 pages. "His memos are the equivalent of a Ph.D. thesis, done over a couple of days," Rashid says, shaking his head.

The memos bring some of the best traits of a university—exploring ideas, conveying them to others and spurring discussion—into the corporate fold. The same features come through in Myhrvold's discussions, whether with freshman employees or with Gates. At one of their weekly sessions, for instance, Myhrvold and Gates engaged in a spirited debate over the merits of so-called generalized sprites, a new twist on an old technique for controlling video frames.

"Intel had a sprite-based strategy," Gates said, his voice rising skeptically. "And we were rather eloquent in telling them that it was a waste of silicon—"

"Yup," Myhrvold interjected.

"But now we're taking a new position?" Gates asked.

"We are," Myhrvold declared. Gates giggled.

"The basic idea behind g-sprites is this," Myhrvold said, snapping forward in his chair and grabbing the memo pad on the nearby coffee table.

During the next few minutes, Myhrvold whizzed through an abbreviated tutorial on computer animation techniques, haphazardly sketching a few rectangles on the paper to bolster a point. Gates picked up on the ideas.

"So, if we're fast enough now to have that guy," pointing to one of the rectangles, "be—I don't know, I'm not pointing at the right thing—"

"Yeah," Myhrvold said, encouragingly.

"—to have this guy be essentially programmable, then that's much better than sprites—"

"—yes—"

"—in the old sense of the word," Gates concluded.

"Right," Myhrvold confirmed. Case closed. Within weeks, Microsoft would gather its energies and begin exploring a new strategy for supporting computer animation. *—Elizabeth Corcoran*

Environmental Change and Violent Conflict

Growing scarcities of renewable resources can contribute to social instability and civil strife

by Thomas F. Homer-Dixon, Jeffrey H. Boutwell and George W. Rathjens

Within the next 50 years, the human population is likely to exceed nine billion, and global economic output may quintuple. Largely as a result of these two trends, scarcities of renewable resources may increase sharply. The total area of highly productive agricultural land will drop, as will the extent of forests and the number of species they sustain. Future generations will also experience the ongoing depletion and degradation of aquifers, rivers and other bodies of water, the decline of fisheries, further stratospheric ozone loss and, perhaps, significant climatic change.

As such environmental problems become more severe, they may precipitate civil or international strife. Some concerned scientists have warned of this prospect for several decades, but the debate has been constrained by lack of carefully compiled evidence. To address this shortfall of data, we assembled a team of 30 researchers to examine a set of specific cases. In studies commissioned by the University of Toronto and the American Academy of Arts and Sciences, these experts reported their initial findings.

The evidence that they gathered points to a disturbing conclusion: scarcities of renewable resources are already contributing to violent conflicts in many parts of the developing world. These conflicts may foreshadow a surge of similar violence in coming decades, particularly in poor countries where shortages of water, forests and, especially, fertile land, coupled with rapidly expanding populations, already cause great hardship.

Before we discuss the findings, it is important to note that the environment is but one variable in a series of political, economic and social factors that can bring about turmoil. Indeed, some skeptics claim that scarcities of renewable resources are merely a minor variable that sometimes links existing political and economic factors to subsequent social conflict.

The evidence we have assembled supports a different view [*see illustration on page 18*]. Such scarcity can be an important force behind changes in the politics and economics governing resource use. It can cause powerful actors to strengthen, in their favor, an inequitable distribution of resources. In addition, ecosystem vulnerability often contributes significantly to shortages of renewable resources. This vulnerability is, in part, a physical given: the depth of upland soils in the tropics, for example, is not a function of human social institutions or behavior. And finally, in many parts of the world, environmental degradation seems to have passed a threshold of irreversibility. In these situations, even if enlightened social change removes the original political, economic and cultural causes of the degradation, it may continue to contribute to social disruption. In other words, once irreversible, environmental degradation becomes an independent variable.

Skeptics often use a different argument. They state that conflict arising from resource scarcity is not particularly interesting, because it has been common throughout human history. We maintain, though, that renewable-resource scarcities of the next 50 years will probably occur with a speed, complexity and magnitude unprecedented in history. Entire countries can now be deforested in a few decades, most of a region's topsoil can disappear in a generation, and acute ozone depletion may take place in as few as 20 years.

Unlike nonrenewable resources—including fossil fuels and iron ore—renewable resources are linked in highly complex, interdependent systems with many

THOMAS F. HOMER-DIXON, JEFFREY H. BOUTWELL and GEORGE W. RATHJENS are co-directors of the project on Environmental Change and Acute Conflict, which is jointly sponsored by the University of Toronto and the American Academy of Arts and Sciences. Homer-Dixon received his Ph.D. in political science from the Massachusetts Institute of Technology in 1989 and is now coordinator of the Peace and Conflict Studies Program at the University of Toronto. Boutwell, who also received his Ph.D. from M.I.T., is associate executive officer and program director of International Security Studies at the American Academy of Arts and Sciences. Rathjens earned his doctorate in chemistry at the University of California, Berkeley, and is currently professor of political science at M.I.T.

ARMY DETACHMENT patrols village in Assam, India, where in 1983 local tribespeople attacked migrant Muslims from Bangladesh. Members of the tribe had long accused the migrants of stealing some of the region's richest farmland. Before troops arrived to restore order, almost 1,700 Bengalis had been massacred in one incident alone.

nonlinear and feedback relations. The overextraction of one resource can lead to multiple, unanticipated environmental problems and sudden scarcities when the system passes critical thresholds.

Our research suggests that the social and political turbulence set in motion by changing environmental conditions will not follow the commonly perceived pattern of scarcity conflicts. There are many examples in the past of one group or nation trying to seize the resources of another. For instance, during World War II, Japan sought to secure oil, minerals and other resources in China and Southeast Asia.

Currently, however, many threatened renewable resources are held in common—including the atmosphere and the oceans—which makes them unlikely to be the object of straightforward clashes. In addition, we have come to understand that scarcities of renewable resources often produce insidious and cumulative social effects, such as population displacement and economic disruption. These events can, in turn, lead to clashes between ethnic groups as well as to civil strife and insurgency. Although such conflicts may not be as conspicuous or dramatic as wars over scarce resources, they may have serious repercussions for the security inter-

ests of the developed and the developing worlds.

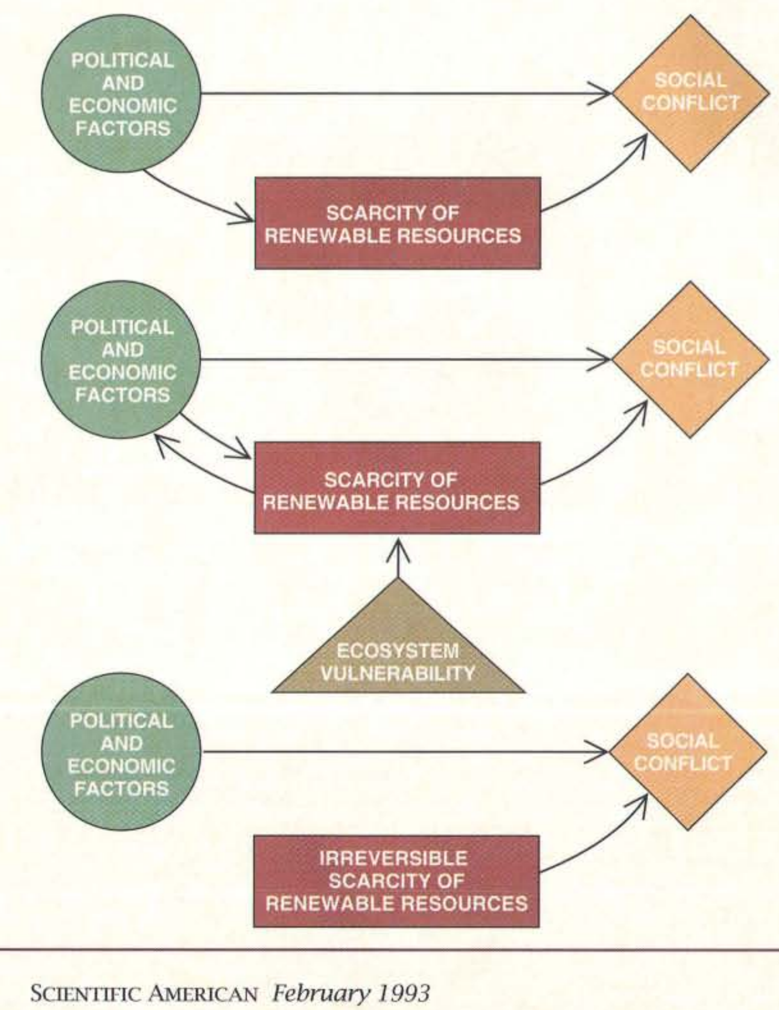
Human actions bring about scarcities of renewable resources in three principal ways. First, people can reduce the quantity or degrade the quality of these resources faster than they are renewed. This phenomenon is often referred to as the consumption of the resource's "capital": the capital generates "income" that can be tapped for human consumption. A sustainable economy can therefore be defined as one that leaves the capital intact and undamaged so that future generations can enjoy undiminished income. Thus, if topsoil creation in a region of farmland is 0.25 millimeter per year, then average soil loss should not exceed that amount.

The second source of scarcity is population growth. Over time, for instance, a given flow of water might have to be divided among a greater number of people. The final cause is change in the distribution of a resource within a society. Such a shift can concentrate supply in the hands of a few, subjecting the rest to extreme scarcity.

These three origins of scarcity can operate singly or in combination. In some cases, population growth by itself will set in motion social stress. Bangladesh, for example, does not suffer from debilitating soil degradation or from the erosion of agricultural land: the annual flooding of the Ganges and Brahmaputra rivers deposits a layer of silt that helps to maintain the fertility of the country's vast floodplains.

But the United Nations predicts that Bangladesh's current population of 120 million will reach 235 million by the year 2025. At about 0.08 hectare per capita, cropland is already desperately scarce. Population density is 785 people per square kilometer (in comparison, population density in the adjacent Indian state of Assam is 284 people per square kilometer). Because all the country's good agricultural land has been exploited, population growth will cut in half the amount of cropland available per capita by 2025. Flooding and inadequate national and community institutions for water control exacerbate the lack of land and the brutal poverty and turmoil it engenders.

Three Views of the Role That Scarcity of Renewable Resources Plays in Violent Conflict



Over the past 40 years, millions of people have migrated from Bangladesh to neighboring areas of India, where the standard of living is often better. Detailed data on the movements are few: the Bangladeshi government is reluctant to admit there is significant migration because the issue has become a major source of friction with India. Nevertheless, one of our researchers, Sanjoy Hazarika, an investigative journalist and reporter at the *New York Times* in New Delhi, pieced together demographic information and experts' estimates. He concludes that Bangladeshi migrants and their descendants have expanded the population of neighboring areas of India by 15 million. (Only one to two million of those people can be attributed to migrations during the 1971 war between India and Pakistan that resulted in the creation of Bangladesh.)

This enormous flux has produced pervasive social changes in the receiving Indian states. Conflict has been triggered by altered land distribution as well as by shifts in the balance of political and economic power between religious and ethnic groups. For instance, members of the Lalung tribe in Assam have long resented Bengali Muslim migrants: they accuse them of stealing the area's richest farmland. In early 1983, during a bitterly contested election for federal offices in the state, violence finally erupted. In the village of Nellie, Lalung tribespeople massacred nearly 1,700 Bengalis in one five-hour rampage.

In the state of Tripura the original Buddhist and Christian inhabitants now make up less than 30 percent of the population. The remaining percentage consists of Hindu migrants from either East Pakistan or Bangladesh. This shift in the ethnic balance precipitated a violent insurgency between 1980 and 1988 that was called off only after the government agreed to return land to dispossessed Tripuris and to stop the influx of Bangladeshis. As the migration has continued, however, this agreement is in jeopardy.

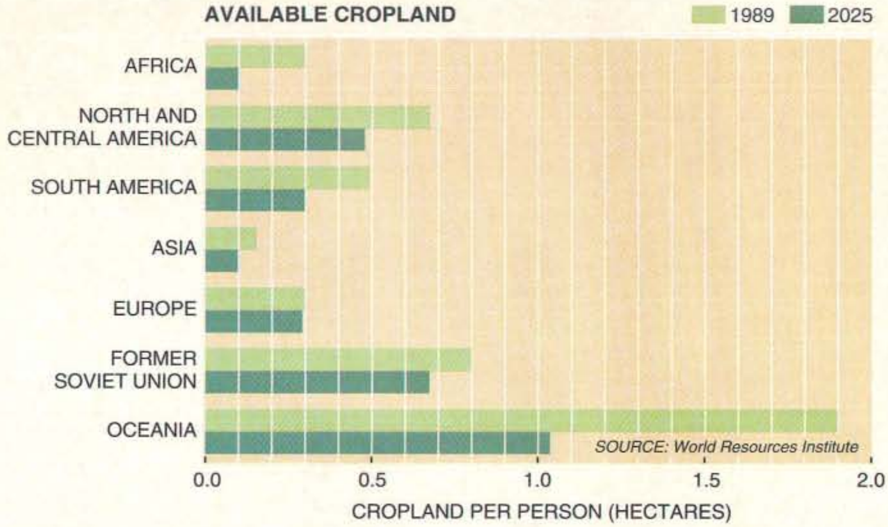
Population movements in this part of South Asia are, of course, hardly new. During the colonial period, the British imported Hindus from Calcutta to administer Assam, and Bengali was made the official language. As a result, the Assamese are particularly sensitive to the loss of political and cultural control in the state. And Indian politicians have often encouraged immigration in

order to garner votes. Yet today changes in population density in Bangladesh are clearly contributing to the exodus. Although the contextual factors of religion and politics are important, they do not obscure the fact that a dearth of land in Bangladesh has been a force behind conflict.

In other parts of the world the three sources of scarcity interact to produce discord. Population growth and reductions in the quality and quantity of renewable resources can lead to large-scale development projects that can alter access to resources. Such a shift may lead to decreased supplies for poorer groups whose claims are violently opposed by powerful elites. A dispute that began in 1989 between Mauritanians and Senegalese in the Senegal River valley, which demarcates the common border between these countries, provides an example of such causality.

Senegal has fairly abundant agricultural land, but much of it suffers from severe wind erosion, loss of nutrients, salinization because of overirrigation and soil compaction caused by the intensification of agriculture. The country has an overall population density of 380 people per square kilometer and a population growth rate of 2.7 percent; in 25 years the population may double. In contrast, except for the Senegal River valley along its southern border and a few oases, Mauritania is for the most part arid desert and semiarid grassland. Its population density is very low, about 20 people per square kilometer, and the growth rate is 2.8 percent a year. The U.N. Food and Agriculture Organization has included both Mauritania and Senegal in its list of countries whose croplands cannot support current or projected populations without a large increase in agricultural inputs, such as fertilizer and irrigation.

Normally, the broad floodplains fringing the Senegal River support productive farming, herding and fishing based on the river's annual floods. During the 1970s, however, the prospect of chronic food shortages and a serious drought encouraged the region's governments to seek international financing for the Manantali Dam on the Bafing River tributary in Mali and for the Diama salt-intrusion barrage near the mouth of the Senegal River between Senegal and Mauritania. The dams were designed to regulate the river's flow for hydropower, to expand irrigated agriculture and to raise water levels in the dry season, permitting year-round barge transport from the Atlantic Ocean to land-locked Mali, which lies to the east of Senegal and Mauritania.

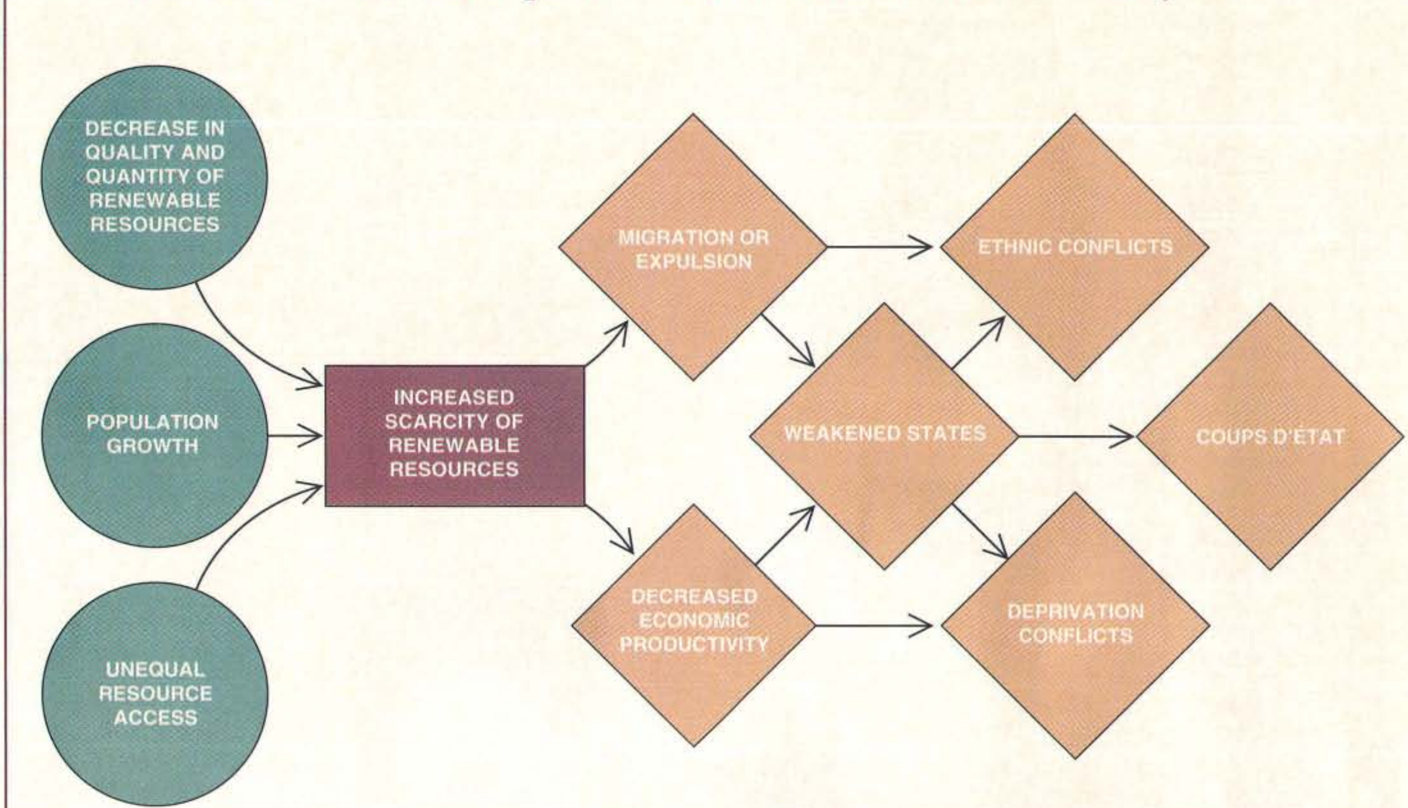


AVAILABLE CROPLAND is expected to decline in many parts of the world by 2025 (*top*) as a result of population growth and the degradation of fertile land. In the Philippines, lack of good land has pushed poor farmers onto steep hillsides (*bottom*). Unterraced farming on such terrain causes severe erosion, which can be seen in the earth-colored gashes on the slopes.

But the plan had unfortunate and unforeseen consequences. As anthropologist Michael M. Horowitz of the State University of New York at Binghamton has shown, anticipation of the new dams raised land values along the river in areas where high-intensity agriculture was to become feasible. The elite in Mauritania, which consists primarily of white Moors, then rewrote legislation governing land ownership, effectively abrogating the rights of black Africans to continue farming, herding and fishing along the Mauritanian riverbank.

There has been a long history of racism by white Moors in Mauritania toward their non-Arab, black compatriots. In the spring of 1989 the killing of Senegalese farmers by Mauritanians in the river basin triggered explosions of ethnic violence in the two countries. In Senegal almost all of the 17,000 shops owned by Moors were destroyed, and their owners were deported to Mauritania. In both countries several hundred people were killed, and the two nations nearly came to war. The Mauritanian regime used this occasion to activate the new land legislation, declaring the black Mauritanians who lived alongside the river to be "Senegalese," thereby stripping them of their citizenship; their property was seized. Some 70,000 of the black Mauritanians were forcibly ex-

Some Sources and Consequences of Renewable Resource Scarcity



pelled to Senegal, from where some launched raids to retrieve expropriated cattle. Diplomatic relations between the two countries have now been restored, but neither has agreed to allow the expelled population to return or to compensate them for their losses.

We see a somewhat different causal process in many parts of the world: unequal access to resources combines with population growth to produce environmental damage. This phenomenon can contribute to economic deprivation that spurs insurgency and rebellion. In the Philippines, Spanish and American colonial policies left behind a grossly inequitable distribution of land. Since the 1960s, the introduction of green revolution technologies has permitted a dramatic increase in lowland production of grain for domestic consumption and of cash crops that has helped pay the country's massive external debt.

This modernization has raised demand for agricultural labor. Unfortunately, though, the gain has been overwhelmed by a population growth rate of 2.5 to 3.0 percent. Combined with the maldistribution of good cropland and an economic crisis in the first half of the 1980s, this growth produced a surge in agricultural unemployment.

With insufficient rural or urban industrialization to absorb excess labor, there has been unrelenting downward pressure on wages. Economically desperate, millions of poor agricultural laborers and landless peasants have migrated to shantytowns in already overburdened cities, such as Manila; millions of others have moved to the least productive—and often most ecologically vulnerable—territories, such as steep hillsides.

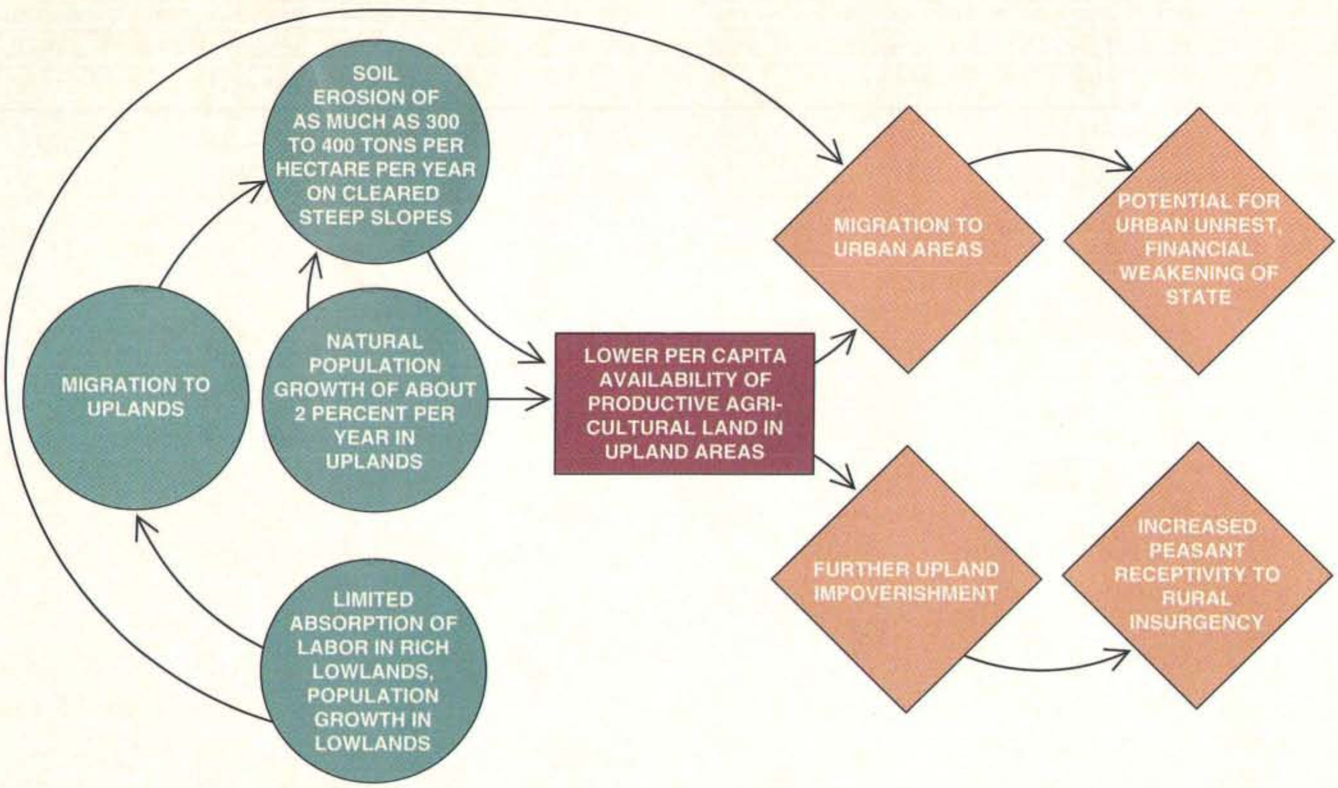
In these uplands, settlers use fire to clear forested or previously logged land. They bring with them little ability to protect the fragile ecosystem. Their small-scale logging, charcoal production and slash-and-burn farming often cause erosion, landslides and changes in hydrologic patterns. This behavior has initiated a cycle of falling food production, the clearing of new plots and further land degradation. Even marginally fertile land is becoming hard to find in many places, and economic conditions are critical for peasants.

The country has suffered from serious internal strife for many decades. But two researchers, Celso R. Roque, the former undersecretary of the environment of the Philippines, and his colleague Maria I. Garcia, conclude that resource scarcity appears to be an increasingly powerful force behind the current communist-led insurgency. The upland struggle—including guerrilla attacks and assaults on military stations—is motivated by the economic deprivation of the landless agricultural laborers and poor farmers displaced into the hills, areas that are largely beyond the control of the central government. During the 1970s and 1980s, the New People's Army and the National Democratic Front found upland peasants receptive to revolutionary ideology, especially where coercive landlords and local governments left them little choice but to rebel or starve. The revolutionaries have built on indigenous beliefs and social structures to help the peasants focus their discontent.

Causal processes similar to those in the Philippines can be seen in many other regions around the planet, including the Himalayas, the Sahel, Indonesia, Brazil and Costa Rica. Population growth and unequal access to good land force huge numbers of people into cities or onto marginal lands. In the latter case, they cause environmental damage and become chronically poor. Eventually these people may be the source of persistent upheaval, or they may migrate yet again, stimulating ethnic conflicts or urban unrest elsewhere.

The short but devastating "Soccer War" in 1969 between El Salvador and Honduras involved just such a combin-

AN EXAMPLE FROM THE PHILIPPINES



ation of factors. As William H. Durham of Stanford University has shown, changes in agriculture and land distribution beginning in the mid-19th century concentrated poor farmers in El Salvador's uplands. Although these peasants developed some understanding of land conservation, their growing numbers on very steep hillsides caused deforestation and erosion. A natural population growth rate of 3.5 percent further reduced land availability, and as a result many people moved to neighboring Honduras. Their eventual expulsion from Honduras precipitated a war in which several thousand people were killed in a few days. Durham notes that the competition for land in El Salvador leading to this conflict was not addressed in the war's aftermath and that it powerfully contributed to the country's subsequent, decade-long civil war.

In South Africa the white regime's past apartheid policies concentrated millions of blacks in the country's least productive and most ecologically sensitive territories. High natural birth rates exacerbated population densities. In 1980 rural areas of the Ciskei homeland supported 82 persons per square kilometer, whereas the surrounding Cape Province had a rural density of two. Homeland residents had, and have, little capital and few skills to manage resources. They remain the victims of corrupt and abusive local governments.

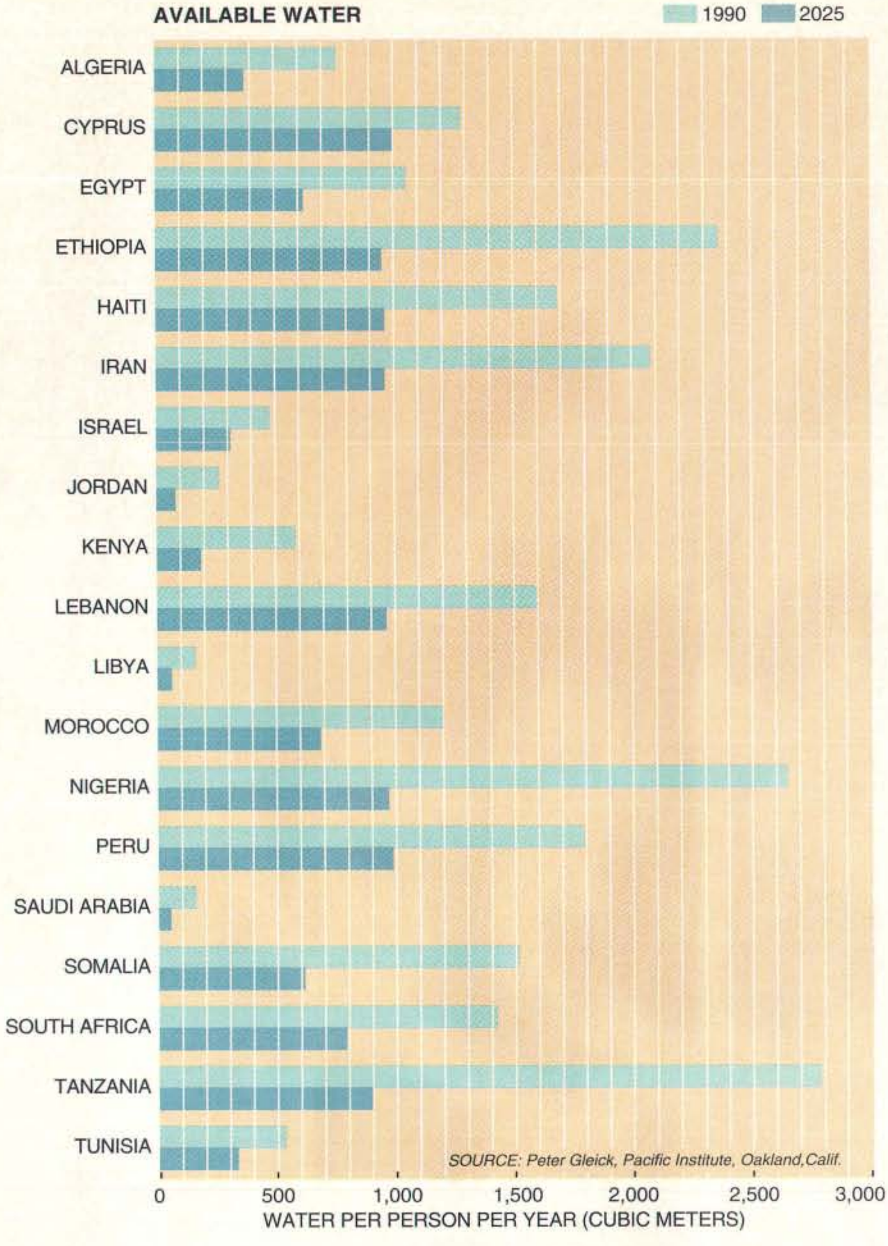
Sustainable development in such a situation is impossible. Wide areas have been completely stripped of trees for fuelwood, grazed down to bare dirt and eroded of topsoil. A 1980 report concluded that nearly 50 percent of Ciskei's land was moderately or severely eroded; close to 40 percent of its pasture was overgrazed. This loss of resources, combined with the lack of alternative employment and the social trauma caused by apartheid, has created a subsistence crisis in the homelands. Thousands of people have migrated to South African cities. The result is the rapid growth of squatter settlements and illegal townships that are rife with discord and that threaten the country's move toward democratic stability.

Dwindling natural resources can weaken the administrative capacity and authority of government, which may create opportunities for violent challenges to the state by political and military opponents. By contributing to rural poverty and rural-urban migration, scarcity of renewable resources expands the number of people needing assistance from the government. In response to growing city populations, states often introduce subsidies that distort prices and cause misallocations of capital, hindering economic productivity.

Simultaneously, the loss of renewable resources can reduce the production of wealth, thereby constraining tax revenues. For some countries, this widening gap between demands on the state and its capabilities may aggravate popular grievances, erode the state's legitimacy and escalate competition between elite factions as they struggle to protect their prerogatives.

Logging for export markets, as in Southeast Asia and West Africa, produces short-term economic gain for parts of the elite and may alleviate external debt. But it also jeopardizes long-term productivity. Forest removal decreases the land's ability to retain water during rainy periods. Flash floods then damage roads, bridges, irrigation systems and other valuable infrastructure. Erosion of hillsides silts up rivers, reducing their navigability and their capacity to generate hydroelectric power. Deforestation can also hinder crop production by altering regional hydrologic cycles and by plugging reservoirs and irrigation channels with silt [see "Accounting for Environmental Assets," by Robert Repetto; SCIENTIFIC AMERICAN, June 1992].

In looking at China, Václav Smil of the University of Manitoba has estimated



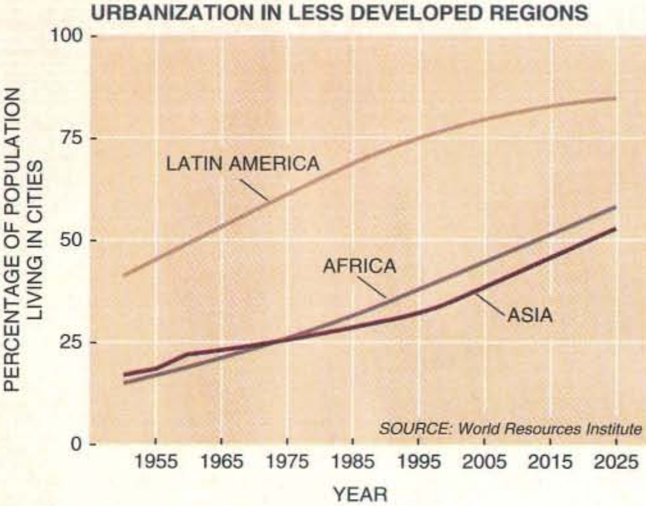
WATER SHORTAGES may be severe in the future. In 2025 several nations (*top*) will have less than 1,000 cubic meters of water per person—the minimum amount considered necessary for an industrialized nation. In Ethiopia, water is already so scarce that some women walk miles to find it and lug it home (*bottom*).

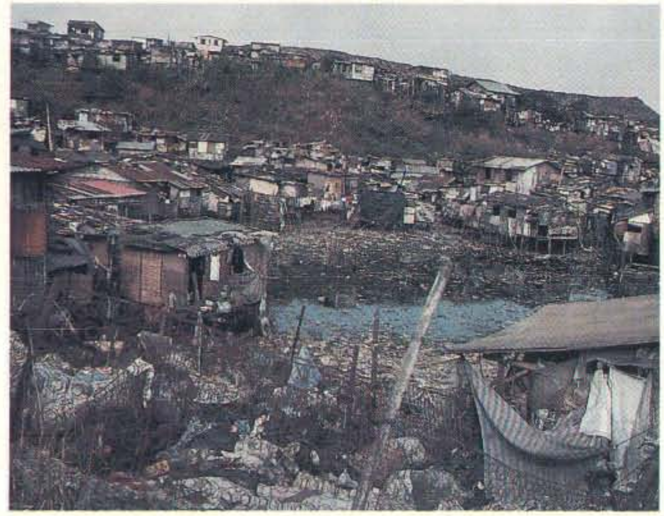
the combined effect of environmental problems on productivity. The main economic burdens he identifies are reduced crop yields caused by water, soil and air pollution; higher human morbidity resulting from air pollution; farmland loss because of construction and erosion; nutrient loss and flooding caused by erosion and deforestation; and timber loss arising from poor harvesting practices. Smil calculates the current annual cost to be at least 15 percent of China's gross domestic product; he is convinced the toll will rise steeply in the coming decades. Smil also estimates that tens of millions of Chinese will try to leave the country's impoverished interior and northern regions—where water and fuelwood are desperately scarce and the land often badly damaged—for the booming coastal cities. He anticipates bitter disputes among these regions over water sharing and migration. Taken together, these economic and political stresses may greatly weaken the Chinese state.

Water shortages in the Middle East will become worse in the future and may also contribute to political discord. Although figures vary, Miriam R. Lowi of Princeton University estimates that the average amount of renewable fresh water available annually to Israel is about 1,950 million cubic meters (mcm). Sixty percent comes from groundwater, the rest from river flow, floodwater and wastewater recycling. Current Israeli demand—including that of settlements in the occupied territories and the Golan Heights—is about 2,200 mcm. The annual deficit of about 200 mcm is met by overpumping aquifers.

As a result, the water table in some parts of Israel and the West Bank has been dropping significantly. This depletion can cause the salinization of wells and the infiltration of seawater from the Mediterranean. At the same time, Israel's population is expected to increase from the present 4.6 million to 6.5 million people in the year 2020, an estimate that does not include immigration from the former Soviet Union. Based on this projected expansion, the country's water demand could exceed 2,600 mcm by 2020.

Two of the three main aquifers on which Israel depends lie for the most part under the West Bank, although their waters drain into Israel. Thus, nearly 40 percent of the groundwater Israel uses originates in occupied territory. To protect this important source, the Israeli government has strictly limited water use on the West Bank. Of the 650 mcm of all forms of water annually available there, Arabs are allowed to use only 125 mcm. Israel restricts the num-





GROWTH OF CITIES, in part a result of increasing rural poverty and of migration, will be dramatic in the developing world (*left*). In Manila the "Smoky Mountains" squatter settlement is home to poor peasants arriving by ship from the provinces (*right*). The Filipinos named the settlement after the perpetually smoldering garbage dump on which it is constructed.

ber of wells Arabs can drill in the territory, the amount of water Arabs are allowed to pump and the times at which they can draw irrigation water.

The differential in water access on the West Bank is marked: on a per capita basis, Jewish settlers consume about four times as much water as Arabs. Arabs are not permitted to drill new wells for agricultural purposes, although Mekorot (the Israeli water company) has drilled more than 30 for settlers. Arab agriculture in the region has suffered because some Arab wells have become saline as a result of deeper Israeli wells drilled nearby. The Israeli water policy, combined with the confiscation of agricultural land for settlers as well as other Israeli restrictions on Palestinian agriculture, has encouraged many West Bank Arabs to abandon farming. Those who have done so have become either unemployed or day laborers within Israel.

The entire Middle East faces increasingly grave and tangled problems of water scarcity, and many experts believe these will affect the region's stability. Concerns over water access contributed to tensions preceding the 1967 Arab-Israeli War; the war gave Israel control over most of the Jordan Basin's water resources. The current Middle East peace talks include multilateral meetings on water rights, motivated by concerns about impending scarcities.

Although "water wars" are possible in the future, they seem unlikely given the preponderance of Israeli military power. More probably, in the context of historical ethnic and political disputes, water shortages will aggravate tensions and unrest within societies in the Jordan River basin. In recent U.S. congressional testimony, Thomas Naff of the University of Pennsylvania noted that "rather than warfare among riparians in the immediate future...what is more likely to ensue from water-related crises in this decade is internal civil disorder, changes in regimes, political radicalization and instability."

Scarcities of renewable resources clearly can contribute to conflict, and the frequency of such unrest will probably grow in the future. Yet some analysts maintain that scarcities are not important in and of themselves. What is important, they contend, is whether people are harmed by them. Human suffering might be avoided if political and economic systems provide the incentives and wherewithal that enable people to alleviate the harmful effects of environmental problems.

Our research has not produced firm evidence for or against this argument. We need to know more about the variables that affect the supply of human ingenuity in response to environmental change. Technical ingenuity is needed for the development of, for example, new agricultural and forestry technologies that compensate for environmental deterioration. Social ingenuity is needed for the creation of institutions that buffer people from the effects of degradation and provide the right incentives for technological innovation.

The role of social ingenuity as a precursor to technical ingenuity is often overlooked. An intricate and stable system of markets, legal regimes, financial agencies and educational and research institutions is a prerequisite for the development and distribution of many technologies—including new grains adapted to dry climates and eroded soils, alternative cooking technologies that compensate for the loss of firewood and water-conservation technologies. Not only are poor countries ill endowed with these social resources, but their ability to create and maintain them will be weakened by the very environmental woes such nations hope to address.

The evidence we have presented here suggests there are significant causal links between scarcities of renewable resources and violence. To prevent such turmoil, nations should put greater emphasis on reducing such scarcities. This means that rich and poor countries alike must cooperate to restrain population growth, to implement a more equitable distribution of wealth within and among their societies, and to provide for sustainable development.

FURTHER READING

SCARCITY AND SURVIVAL IN CENTRAL AMERICA: ECOLOGICAL ORIGINS OF THE SOCCER WAR. William H. Durham. Stanford University Press, 1979.

ON THE POLITICAL CONSEQUENCES OF SCARCITY AND ECONOMIC DECLINE. Ted Gurr in *The International Studies Quarterly*, Vol. 29, No. 1, pages 51–75; March 1985.

ENVIRONMENT AND THE POOR: DEVELOPMENT STRATEGIES FOR A COMMON AGENDA. Edited by Jeffrey H. Leonard. Transaction Publishers, 1989.

ON THE THRESHOLD: ENVIRONMENTAL CHANGES AS CAUSES OF ACUTE CONFLICT. Thomas F. Homer-Dixon in *International Security*, Vol. 16, No. 2, pages 76–116; Fall 1991.

Resistance in High-Temperature Superconductors

Researchers are beginning to see how the motion of magnetic vortices in these materials can interfere with the flow of current

by David J. Bishop, Peter L. Gammel and David A. Huse

Levitating trains and high-capacity devices for storing electrical energy were among the many bold visions some physicists entertained after the discovery of high-temperature superconductors in 1986. But several difficulties quickly emerged to temper the promise extended by the ability of these ceramic materials to conduct electricity at high temperatures without resistance. One of the most vexing hindrances has been the destruction of the superconducting state when the material is placed in a magnetic field—a condition crucial for, or at least inescapable in, many envisaged applications. Resistance to current flow can happen when the external magnetic field penetrates the superconductor in the form of discrete bundles called flux lines. Because a line of flux consists of whirlpools of electric current, it is often called a vortex. If these vortices move, they can impede the flow of electrons. Knowing how these vortices move and arrange themselves under various temperature and magnetic-field conditions will be critical in controlling the phenomenon and in maintaining the supercurrent flow.

Fortunately, recent studies have greatly enhanced our knowledge of vortices. Investigators have found that the magnetic-field behavior of superconductors is much richer than formerly thought. Indeed, the vortices have been found to be capable of forming a number of exotic new phases of matter within the family of high-temperature superconductors. To describe these phases—vortex solids, liquids and glasses—workers have been forced to discard some previously held views in superconductivity and to form fresh hypotheses based on modern concepts in condensed-matter physics. To test the new ideas, investigators have devised experimental techniques of unprecedented sensitivity. The work may ultimately point the way to full understanding and, perhaps, to effective use of these new materials.

DAVID J. BISHOP, PETER L. GAMMEL and DAVID A. HUSE are members of the technical staff at AT&T Bell Laboratories in Murray Hill, N.J. Their overlapping tastes in physics were developed under the tutelage of John D. Reppy and Michael E. Fisher at Cornell University, where all three received their doctorates. Bishop, who holds a B.S. from Syracuse University, is a department head at AT&T. His current research interests include the statics and dynamics of magnetic vortices in exotic superconductors, and his outside pursuits include sailing and collecting antiquarian books. Gammel, who earned two bachelor degrees from the Massachusetts Institute of Technology, has investigated single-charge transport in small tunnel junctions. He also works on his violin playing and vegetable gardening. Huse has been primarily interested in the theory of phase transitions in various materials, including spin glasses. He received his B.S. from the University of Massachusetts at Amherst. They write that "none of us races sports cars or flies jet-fighter planes on weekends. For us, a thrill is a good referee report."

In retrospect, one should not be surprised that the knowledge of the superconducting state gathered before 1986 was inadequate to describe high-temperature superconductivity. The early ideas evolved from observations of conventional superconductors. Such materials, generally familiar metals and alloys, conduct electricity without resistance only when cooled to temperatures within a few degrees of absolute zero. In fact, curiosity about the behavior of matter at low temperatures had led the Dutch physicist Heike Kamerlingh Onnes to discover superconductivity in 1911. The finding came about because Onnes had accomplished the experimentally daunting task of liquefying helium, the last of the inert gases to be condensed. Liquid helium enabled Onnes to cool down materials to temperatures near one kelvin of absolute zero. (Absolute zero is equal to −458 degrees Fahrenheit or −273 degrees Celsius.)

According to a perhaps apocryphal story, the finding emerged when Onnes asked a student to measure the electrical resistance of mercury. The student reported that the resistance disappeared when the temperature of the sample fell to 4.2 kelvins. Onnes sent him back to the laboratory to find what Onnes thought was an "error" producing an experimental artifact. After several tries, the error could not be found, and the workers realized they had made a historic discovery. Onnes went on to win the 1913 Nobel Prize in Physics for this and many other important discoveries in low-temperature physics.

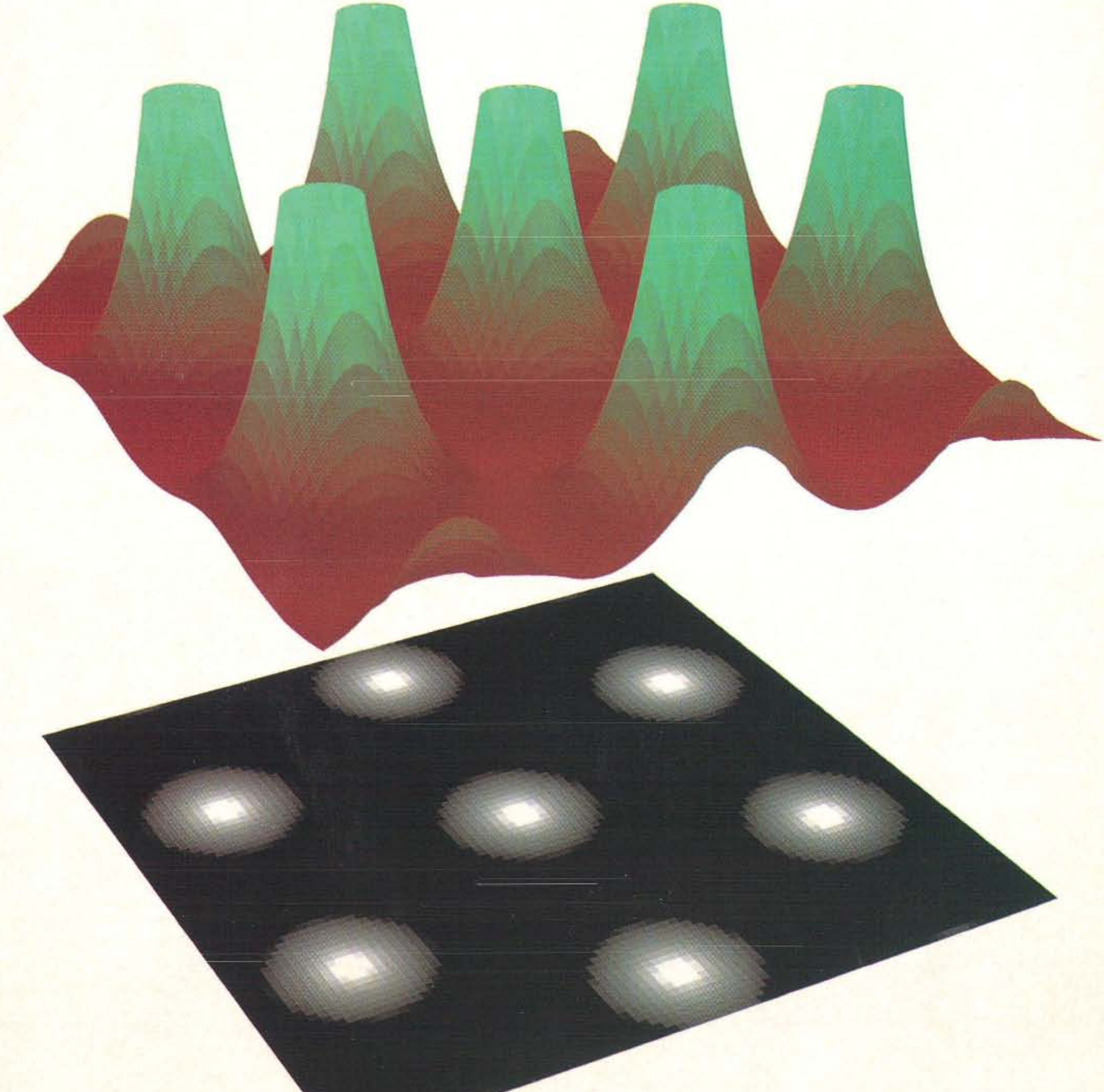
Zero resistance to current flow was not the only reason for amazement. The behavior of superconductors in a magnetic field proved equally astounding. In 1933 two German physicists, Walther Meissner and Robert Ochsenfeld, found that a superconductor can

VORTICES, represented here as green-and-red volcanolike tubes, are discrete bundles of magnetic-field lines that pierce a superconductor. The computer image represents the strength of the magnetic field (plotted as the height of the tubes) across the surface of the sample. The field is largest at the center of each vortex. The projection below the image depicts the vortices as white dots and shows that they form a regular, triangular pattern within the body of the superconductor.

expel magnetic fields when cooled below its superconducting transition temperature. The complete expulsion of a magnetic field is now known as the Meissner effect. Along with the absence of resistance, the ability to exclude magnetic fields propels the enormous research interest in superconductivity.

At this juncture, observation had far outrun theory. The quantum models developed in the 1930s could account for the conductivity in normal metals, but they could not explain the superconducting state. The problem proved particularly intractable; workers did not achieve significant theoretical understanding of the microscopic origins of superconductivity until the 1950s. Then, two Russians, Vitaly L. Ginzburg and Lev D. Landau, developed a phenomenological theory. By looking at what happens during the transition from the normal state to the superconducting one, the scientists were able to formulate a series of equations that could describe the phenomenon. They could not, however, explain why it occurred.

In 1957 John Bardeen, Leon N. Cooper and J. Robert Schrieffer developed



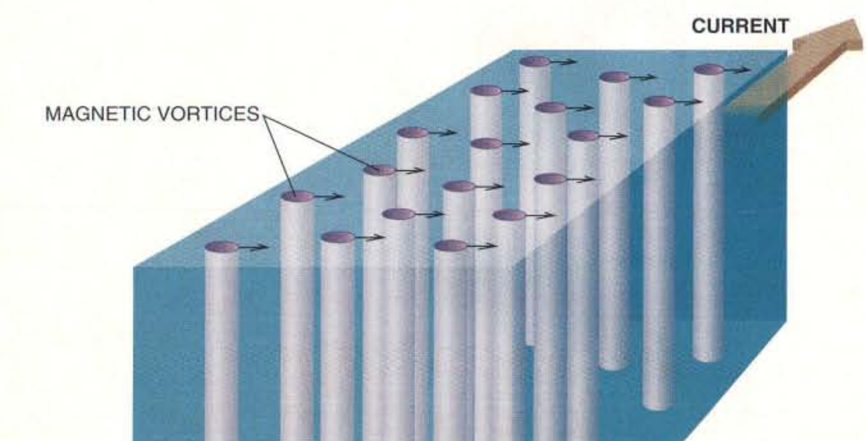
the theory that provided the microscopic explanation for superconductivity. According to the so-called BCS theory, the conduction electrons travel without meeting resistance because they move in pairs, known as Cooper pairs. Electrons form Cooper pairs because they interact with phonons, mechanical vibrations in the crystalline lattice of the metal that resemble sound waves. The movement of the atoms in the lattice tends to neutralize the repulsion that electrons normally have for one another. In fact, it actually produces a small attractive force between electrons. The effectiveness of this interaction depends sharply on temperature. Indeed, the point on a thermal scale at which superconductivity appears is called the transition temperature. At temperatures above this critical point, thermal fluctuations destroy the Cooper pairs and, consequently, the superconductivity of the metal.

The pairing interaction determines two important microscopic distance scales in a superconductor. The first of these is the spatial separation of the electrons in a Cooper pair. This distance is referred to as the coherence length. It is the smallest length in a superconductor over which electronic properties, such as the local resistivity, can change. In typical superconductors the coherence length ranges from hundreds to thousands of angstroms. (These scales of distance are related to atomic reality and so can can be difficult to grasp intuitively—one angstrom equals 10^{-10} meter. The atoms in most materials are spaced one to three angstroms apart.)

The second microscopic characteristic length is related to the strength of the Meissner effect—that is, the ability of a superconductor to expel an applied magnetic field. The effect occurs when a small magnetic field is applied to a superconductor, creating currents that flow near the surface of the material. These induced currents create a magnetic field that precisely cancels the applied field in the rest of the material. The magnitude of these induced currents falls off exponentially with increasing distance from the surface of the superconductor. The length over which this decay occurs is called the magnetic penetration depth. This depth is the shortest distance over which the magnetic field can change in a superconductor. In typical superconductors, this length can vary from hundreds up to tens of thousands of angstroms.

These microscopic lengths define two broadly different categories of superconductors. In type I superconductors the coherence length is longer than the penetration depth. These materials tend to be low-temperature, low-field superconductors. If the field reaches a critical strength (which varies from substance to substance), it enters the material, destroying the superconducting state. Because their lack of resistance disappears at relatively low fields, type I superconductors have little potential for applications or interesting technologies.

Type II superconductors are much more useful. The penetration depth of such superconductors is longer than the coherence length. As a result, they remain superconducting even after the magnetic field enters. Type II superconductors can withstand high fields—up to what is called the upper critical field—and thus can carry the largest currents. All the technologically interesting superconductors, including the known high-temperature materials, are of this type.



CURRENT FLOW through a superconductor (*blue rectangular box*) can be disrupted by vortices (*cylinders*). Each vortex consists of a ring of circulating current induced by an external magnetic field (*not shown*). The applied current adds to the circulating current on one side of the vortex but subtracts from the other. The net result is a force that pushes the vortices at right angles to the current flow; the movement dissipates energy and produces resistance.

In the 1950s the Russian physicist Alexei A. Abrikosov published the basic theory of how a conventional type II superconductor behaves in a magnetic field. Building on the work of Ginzburg and Landau, he showed that the magnetic response of a type II superconductor below the critical temperature depends on the strength of the applied field and on the temperature. Such a relation can be represented by a magnetic phase diagram [*see illustration on page 28*], which shows that a conventional superconductor has three distinct magnetic states.

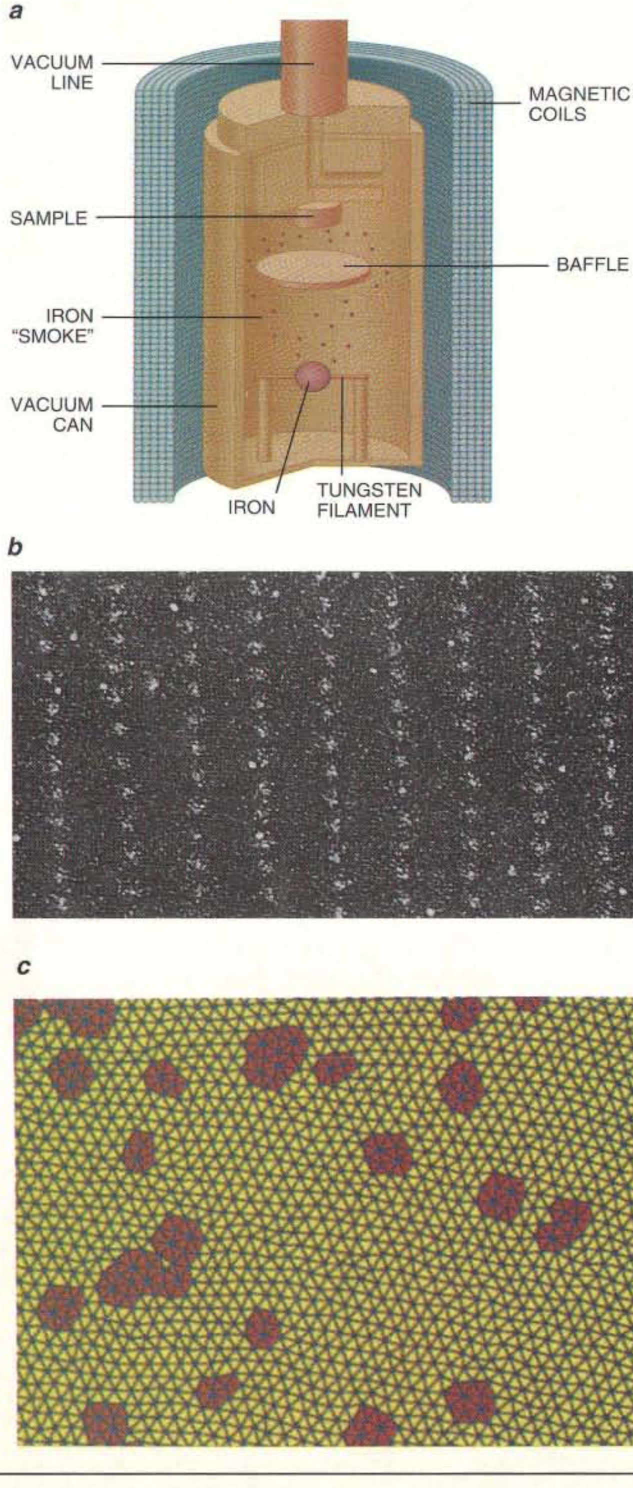
The first one is simply the Meissner state—that is, the state in which the material fully expels the applied field. The superconductor exists in this state as long as the applied magnetic field remains below a certain strength. This field, called the lower critical field, in general depends on temperature.

The second state emerges if the applied field increases to a value higher than the lower critical field. At this point, the magnetic field can still penetrate the superconductor but not completely or uniformly. Instead discrete flux lines, forming tubular intrusions of the applied field, pierce the sample. The quantum mechanics of the superconductor requires that each flux line have exactly the same magnitude. This unit of flux is known as the flux quantum. Because each flux line must have the same strength, any change in the applied magnetic field must change the density of the flux lines. In other words, as the field varies, the distance between the lines changes in response. The mini-

Visualizing the Superconducting Flux Lattice

As children, we all "decorated" the magnetic-field lines of a permanent magnet by using a piece of paper and iron filings. Some of us are still doing it. Specifically, we can decorate the magnetic field that can permeate a superconductor. A small magnetic field enters the superconductor in discrete bundles called flux or vortex lines. The lines arrange themselves in a regular pattern. Several techniques, including neutron scattering and scanning tunneling microscopy, can reveal the pattern, but magnetic decoration is perhaps the simplest and most direct.

The decoration apparatus (*a*), about 10 centimeters high and three centimeters in diameter, consists of only a few key components. The superconductor to be studied rests inside a vacuum can filled with helium gas. We apply a magnetic field with the coils and cool the sample to below its transition temperature. We then heat up the tungsten filament, which has a blob of iron attached to it. The iron particles evaporate. The helium gas in the can cools the iron particles, producing a slowly drifting magnetic "smoke." The iron particles in the smoke are quite small, about 100 angstroms in diameter. They drift around the baffle, which protects the sample from the heat, to the surface of the superconductor. There they decorate the regions where the magnetic-flux lines pass through the surface. The iron particles "stick" to the surface because of the slight attractive forces that exist between all particles. This attraction, called the van der Waals force, acts as an "atomic glue." The sample can warm up to room temperature and still retain the iron particles. We can then use an electron microscope to form a direct picture of the iron particles, which replicates the original flux lattice pattern.

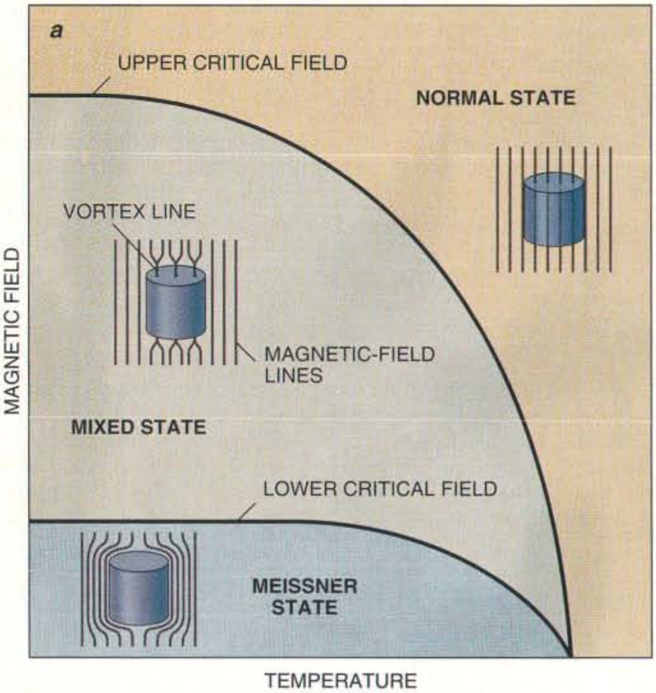


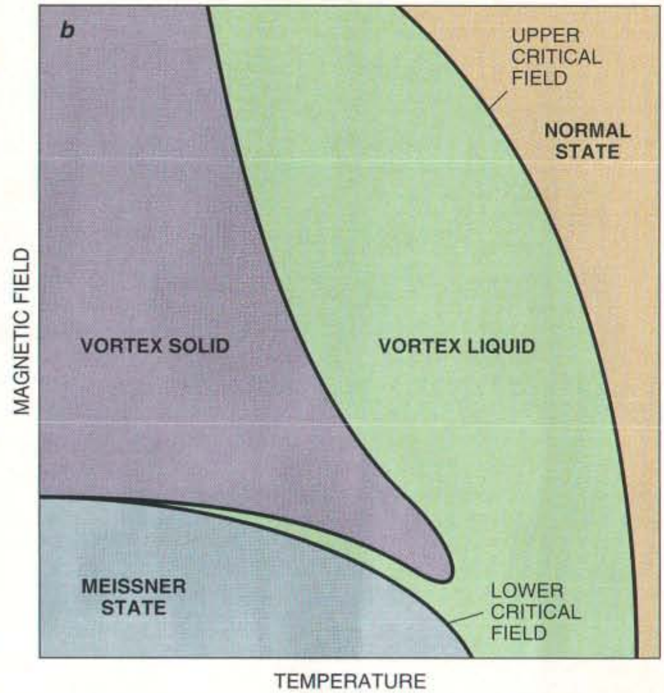
The flux lines appear as dots, revealing the well-ordered nature of the lattice (a photograph of the vortex lattice is on page 29). From such pictures, investigators can determine the amount of magnetic flux per flux line. This amount is a fundamental constant for superconductors known as the flux quantum, Φ_0. For all known superconductors, $\Phi_0 = hc/2e$, where h is Planck's constant, c is the speed of light and e is the charge on the electron. The "2" in the denominator is a direct consequence of the fact that the electrons in superconductors travel in pairs. In the early days of high-temperature superconductivity, some researchers thought the flux quantum might have a different value for these materials. Experiments such as these, which simply count the number of flux lines, quickly ruled out that possibility. By counting, one can show that the ratio of the applied magnetic field to the density of flux lines is equal to the flux quantum.

Decoration experiments have enabled us to see many other novel structures. The pattern of the flux lines will be different if the applied magnetic field strikes the sample at an angle with respect to a major crystallographic axis. Instead of a regular lattice, flux chains appear (*b*).

Several kinds of quantitative analysis are possible with such images. After the locations of the flux lines are digitized, a computer can draw in lines between all of the points in the flux lattice (*c*). In a perfect triangular lattice, each flux line has six nearest neighbors. The defects in the lattice appear as flux lines with different numbers of nearest neighbors. The defects have been shaded red.

Such decorations show that the superconducting flux lattice can take on a specific type of pattern called hexatic order. In such ordered structures the positions of the particles can be random, but the bond angles between nearest neighbors are similar. For the triangulation pattern shown, the bond angles are roughly the same from one end of the picture to the other. But because of the defects, the particles are spaced evenly only over short distances. The easiest way to see this bond-angle order is to place the edge of the page near your eye and to sight along the rows.





MAGNETIC PHASE DIAGRAMS show what happens when a type II superconductor is immersed in a magnetic field. Diagram *a* depicts the three phases present in a conventional superconductor. In the Meissner state (*lower left*) the applied field is expelled. In the mixed (or vortex solid) state the field penetrates in discrete bundles, or vortex lines. In the normal state the field destroys superconductivity and penetrates the material uniformly. Diagram *b* shows that high-temperature superconductors have similar phases, except for a vortex liquid regime. This state exists because thermal fluctuations melt the vortex solid, which is either a lattice (for clean superconductors) or a glass (for dirty ones).

mum energy configuration for such an array of flux lines (as seen from a bird's-eye view above the surface) is a triangular lattice.

The structure of an individual flux line depends on the coherence length and penetration depth. Each line has a small core. The diameter of the core depends on the coherence length. Inside the core, the material is a normal metal. Circulating around the core are supercurrents. (This circulating current is the reason physicists call the core a vortex line.) These supercurrents produce a magnetic field, and the distance over which this magnetic field persists is the magnetic penetration depth. Researchers can make the vortices visible by using small magnetic particles [*see box on preceding page*]. In such images the very well ordered triangular lattice becomes apparent.

The third and final magnetic state of a superconductor emerges if the applied field reaches a second, higher critical point. Above this upper critical field, the superconductivity is completely destroyed, restoring the material to its normal state. The destruction occurs because increases in the strength of the magnetic field force the vortex lines closer together. When the vortex cores, which behave as normal metals, overlap too much, there is no longer enough room between the vortices to maintain superconductivity.

The descriptions of the three magnetic states seemed to detail sufficiently well the effects of an applied magnetic field on superconductors. Then, in 1986, J. Georg Bednorz and K. Alex Müller of the IBM Zürich Research Center came across a new class of type II superconductor. These materials, a family of copper oxide ceramics, were found in some cases to superconduct at a temperature that exceeded 120 kelvins. In contrast, the highest critical temperatures for conventional superconductors lie in the range of 20 to 25 kelvins. The high-temperature superconductors galvanized the scientific world because the materials could easily be cooled with liquid nitrogen, which in bulk costs less than 10 cents per liter (compared to $5 a liter for liquid helium). Even small laboratory-grade refrigerators can cool below the transition temperature of the new superconductors.

As exciting as the high critical temperatures were, a disturbing fact came to light when their properties were studied as a function of an applied magnetic field. Specifically, the high-temperature superconductors did not conform to Abrikosov's successful model. The discrepancies were discovered when investigators studied the materials in magnetic fields that would be necessary in technological use. The strengths of the fields range up to about 10 teslas (a tesla is roughly 20,000 times the strength of the earth's magnetic field). In these fields the resistance of some of these materials did not fall below that of ordinary copper wire until the temperature dropped to only 20 to 30 percent of the superconducting transition temperature. In certain cases, the resistance of some materials in a field remained 100 times higher than that of copper. The advantages of a high-temperature superconductor seemed lost. Additional experiments uncovered the reason. The vortex lines were behaving in an unusual way: they were not always arranging themselves in a rigid, triangular lattice. Instead researchers found that the vortex lattice could "melt" into a liquidlike state. This behavior was suppressing the material's transition to superconductivity.

There are a variety of reasons why this novel state of matter, a vortex liquid, should hinder current flow in the high-temperature superconductors. Perhaps the most convenient way to understand the effect is to imagine vortex lines in a superconductor as rubber bands. Vortex lines and rubber bands

tend to stay short, because making a line longer or stretching a rubber band costs energy. Thermal fluctuations, however, oppose that tendency. Such fluctuations make the atoms in a solid and the vortex lines vibrate with a larger amplitude as the temperature rises. The vortex lines then "stretch." The energy in a vortex line tries to restore the line to its unstretched state.

This restoring force is a function of the coherence length and penetration depth. Long coherence lengths or short penetration depths produce a good deal of restoring force and limit the thermal vibrations of the vortex lines. Most ordinary type II superconductors have such characteristics. The restoring force dominates, keeping the lines straight and short. Thus, thermal fluctuations of the vortex lines are small.

On the other hand, high-temperature superconductors have virtually the opposite characteristics: the coherence lengths are short and the penetration depths long. The coherence length is sometimes as short as a few angstroms, which is about 10 to 100 times below that of conventional superconductors. The penetration depth of high-temperature superconductors ranges from 1,000 to more than 100,000 angstroms; the values exceed that of many conventional superconductors by a factor of 10 to 100.

Coupled with the high transition temperatures, the extreme values of the coherence lengths and penetration depths mean that large thermal fluctuations of the vortex lines occur in the high-temperature superconductors. Indeed, at sufficiently high temperatures the lines vibrate enough to "melt" the vortex lattice. The phenomenon is similar to the way that the thermal vibrations of water molecules can cause ice to melt into water. For some high-temperature materials, the vortex liquid state persists over a temperature range wider than that of the lattice state.

Why does a vortex liquid affect the resistance of the superconductor? The answer lies in thinking about what happens when a current is sent through a type II superconductor in an applied magnetic field. Recall that each vortex line consists of flowing currents circulating around a normal (nonsuperconducting) core. When an applied current flows through the sample, it adds to the circulating current on one side of the vortex and subtracts from it on the other side. As a result, a force acts on the vortex line. The force tends to make the vortex move in a direction at right angles to both the vortex line and the applied current. This force is the Magnus force. It is similar to the lift generated by an airplane wing, a situation in which air flows faster over the upper surface of the wing than it does over the lower surface. If vortex lines move in response to the Magnus force, they will dissipate the energy in the flowing current. Specifically, the dissipation induces a voltage and thus resistance in a sample.

Measurement of this resistance shows how the vortex liquid behaves like ordinary water near the melting point. We have explored the resistance of a very clean piece of the high-temperature superconductor $YBa_2Cu_3O_7$ (yttrium-barium2-copper3-oxygen7, which is often shortened to YBCO, pronounced "ibco") as a function of temperature in a fixed magnetic field. At high temperatures (that is, in the vortex liquid phase), the resistance indicated by the data is high. Lowering the temperature froze the vortex liquid into the vortex lattice state. Hence, the lines were no longer free to move, and the resistance disappeared.

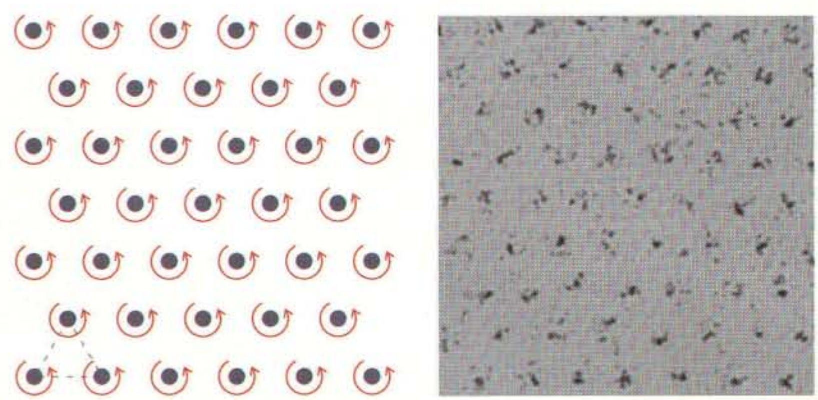
Such resistance measurements also showed that the vortex liquid is slightly supercooled before it freezes. The phenomenon resembles what one finds in clean water, where the liquid phase can to some extent persist below the freezing point. Supercooling can be expressed more technically: the behavior of the substance on heating does not precisely retrace that found on cooling [*see illustration on page 31*]. These processes are said to be hysteretic.

Yet insight into how the vortex liquid state behaves and freezes into a lattice leaves open a question essential for applications. The vortex liquid freezes into a regular lattice only if the material is clean. But what happens when the superconductor is "dirty"—that is, if chemical impurities and defects reside in the atomic lattice?

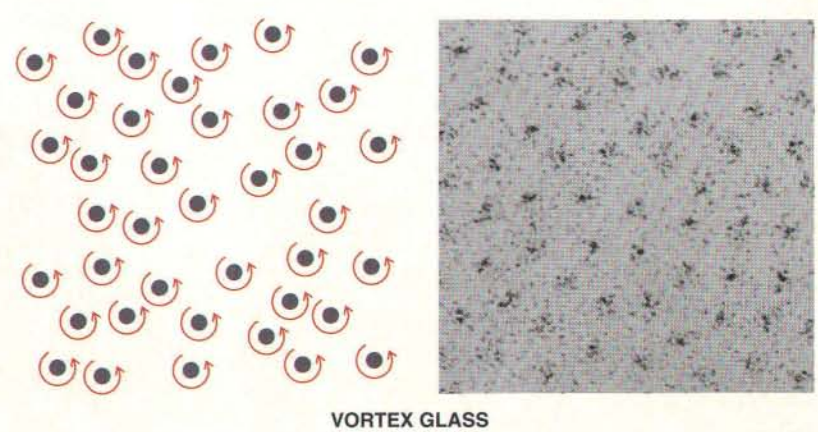
The question is not trivial. Supercon-

States of a Vortex Solid

A superconductor in a magnetic field "freezes" solid in two ways. If the material is clean, the vortex lines will fall into a regular triangular array, forming a vortex lattice. If the substance has many defects or impurities, the lines will develop a disordered pattern, forming a vortex glass.



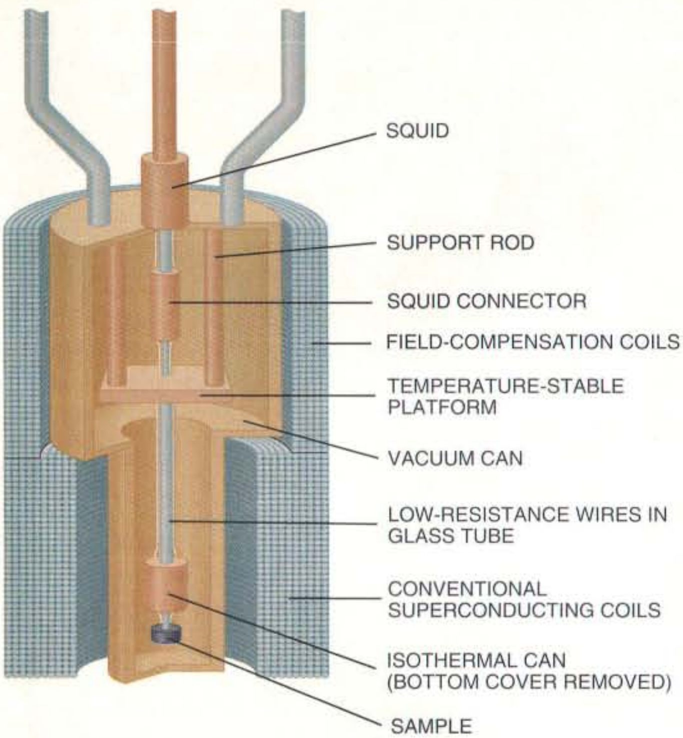
VORTEX LATTICE

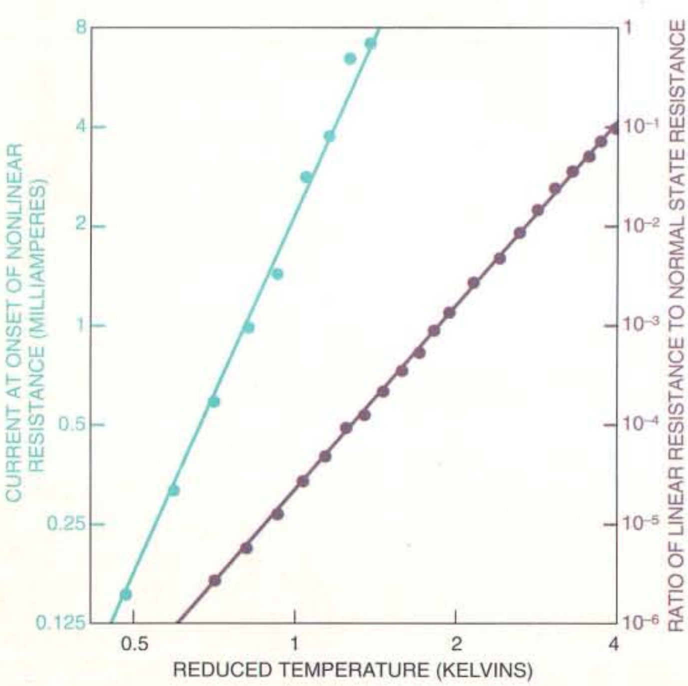


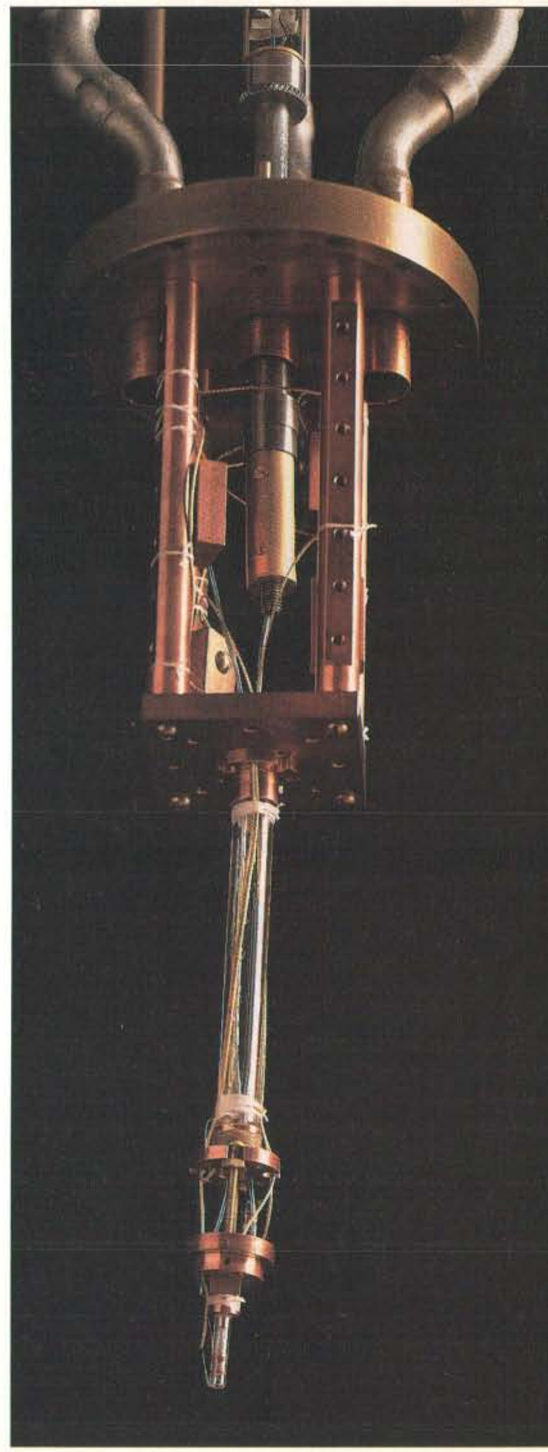
VORTEX GLASS

The SQUID Picovoltmeter

The device probes the different vortex states. The diagram (*top left*) shows the essential features of the picovoltmeter. The conventional superconducting and field-compensation coils apply to the sample a field of up to seven teslas. The isothermal can keeps the temperature of the sample to within a few millikelvins. Low-resistance wires running through a glass tube connect the sample to the SQUID, which measures minute electrical changes. In the photograph (*right*) the magnetic coils and the vacuum and isothermal cans have been removed for clarity. Measurements conducted with the device have confirmed the vortex glass model. One experiment, the results of which are displayed (*bottom left*), looked at the current (*blue*) and resistance (*purple*) in a region where the electrical properties of the sample are nonlinear. The data lie in a straight line, as predicted by theory. The reduced temperature is the difference between the temperature of the sample and that where the superconducting vortex glass phase first occurs.





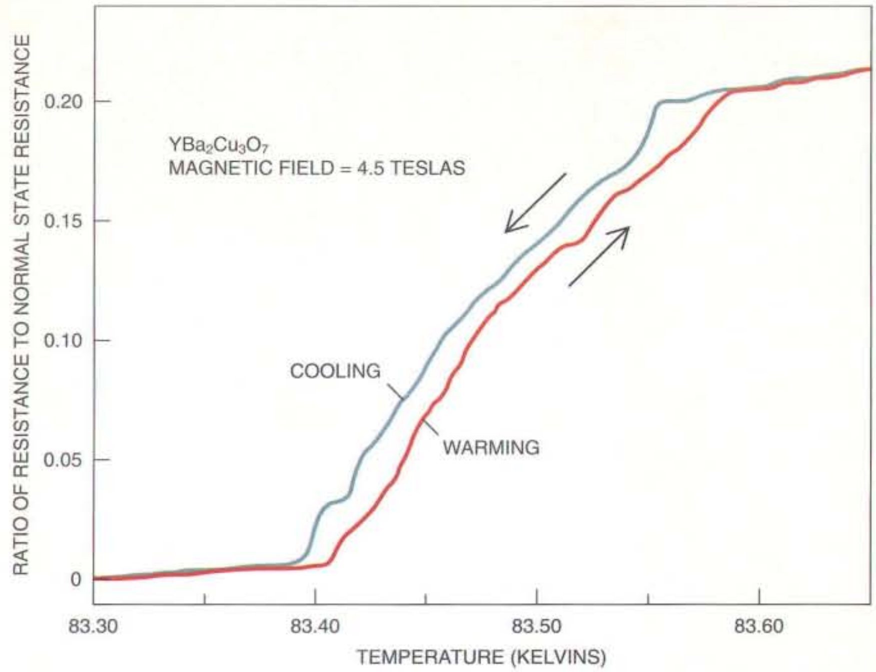


ductors envisioned for technological use must inevitably be dirty. In fact, researchers working with conventional superconductors carefully engineer such defects into the material. Generally, the dirtier a superconductor is, the more current it can carry. Such imperfections are desirable because they "pin" vortices and prevent them from moving in response to the Magnus force. Vortex lines prefer to sit at pinning sites in the crystal lattice because in doing so they lower their energy. The situation is analogous to that of a marble rolling around on top of a table that contains a few small holes. Common experience tells us that the marble prefers to sit in one of the holes in the table, where its gravitational potential energy is lowest.

Pinning has a characteristic effect on the vortex solid in a superconductor: it disrupts the regular lattice pattern that would otherwise form in an ideal, pure material. In other words, the pinning prevents the material from condensing into a perfect vortex solid in strong magnetic fields. The phase that forms instead is what researchers now describe as a vortex glass. The term is appropriate because the positions of the vortices form an irregular, disordered pattern, similar to that assumed by molecules in glass.

The vortex glass idea was not widely accepted when it was first proposed in 1989. Other descriptions, such as those that treat the vortex lines as individual particles, could also account for the observed behavior of the high-temperature materials. The vortex glass model, however, made several testable predictions. It postulated that, given a sufficiently large concentration of pinning defects, the vortex liquid would freeze smoothly into a glass. This behavior contrasts with that shown for pure materials, in which the vortex liquid solidifies rather abruptly and in a hysteretic manner. The vortex glass model also described the behavior of the resistivity as a function of temperature, current and magnetic field.

A clear verification of the vortex glass model came about only when researchers could carry out extremely sensitive transport measurements of a type not usually done in superconductors. Specifically, experimenters designed an apparatus that could measure the voltage across a high-temperature superconductor with subpicovolt (10^{-12} volt) resolution—an accuracy previously unavailable. The picovoltmeter used a superconducting quantum interference device, or SQUID. Such devices rely on quantum effects to measure minute current and voltage changes. With a SQUID, the picovoltmeter had a sensitivity about one million times greater



COOLING AND HEATING of a very clean crystal of the superconductor YBCO in a magnetic field produce resistance plots that do not exactly retrace one another. The measurement shows that the vortex lattice melts abruptly. In effect, the vortex liquid can be slightly "supercooled" before it freezes, much as pure water can.

than that of an ordinary voltmeter. The resolution was sufficiently high to confirm or dispute the vortex glass theory.

The principles behind the picovoltmeter itself are rather simple. Samples are placed in an insulating container that can maintain the temperature inside to within a few millikelvins. Superconducting coils surround the container and apply a uniform magnetic field to the sample. Current is sent through wires connected to the sample, and the SQUID then measures the resistance of the sample. The SQUID and superconducting magnets are conventional, low-temperature superconductors—an example of the old technology helping us to measure and grasp the new.

The apparatus resoundingly confirmed the predictions of the vortex glass model. The measured resistances and currents matched those predicted by the model, smoothly going to zero as the temperature was reduced to the freezing point of the liquid [*see box on opposite page*]. This smooth behavior is very different from that found for very clean crystals: in them, the phase transition is sudden and hysteretic. The observation shows the importance of pinning-induced disorder—the role of "dirt," so to speak—in changing the dynamics of the melting transition. Instead of a solid, the vortex liquid in the disordered crystal freezes into a vortex glass.

The high-temperature superconductors have proved to be a wonderful testing ground for our knowledge of type II superconductivity. For instance, we can now conclude that the vortex glass also exists in conventional superconductors, although the state may be hard to see. Nevertheless, it remains to be seen whether the knowledge can be translated effectively into applications. Researchers are actively looking for the kind of defects that could pin vortices most effectively. Much has been accomplished to fashion superconducting wires and to improve their current-carrying capability. Our present microscopic understanding of the various vortex states can only help us engineer better materials for the applications we all so eagerly await.

FURTHER READING

SUPERCONDUCTIVITY. Edited by R. D. Parks. Marcel Dekker, 1969.

INTRODUCTION TO SUPERCONDUCTIVITY. Michael Tinkham. Robert E. Krieger Publishing, 1980.

MAGNETIC FLUX-LINE LATTICES AND VORTICES IN THE COPPER OXIDE SUPERCONDUCTORS. D. J. Bishop, P. L. Gammel, D. A. Huse and C. A. Murray in *Science*, Vol. 255, pages 165–172; January 10, 1992.

ARE SUPERCONDUCTORS REALLY SUPERCONDUCTING? D. A. Huse, Matthew P. A. Fisher and Daniel S. Fisher in *Nature*, Vol. 358, No. 6387, pages 553–559; August 13, 1992.

Zinc Fingers

They play a key part in regulating the activity of genes in many species, from yeast to humans. Fewer than 10 years ago no one knew they existed

by Daniela Rhodes and Aaron Klug

One of the most fascinating questions in biology today asks how genes are turned on in multicellular organisms. If a gene is to be activated, several proteins known as transcription factors must attach themselves to a segment of the gene called the promoter. This assembly forms a kind of "on switch": it enables an enzyme to transcribe a second genetic segment from DNA into RNA. In most cases, the resulting RNA molecule serves as a template for synthesis of a specific protein, or string of amino acids; sometimes RNA itself is the final product. Yet scientists have continued to wonder exactly how a transcription factor picks out its particular docking site on a promoter, distinguishing that site from the masses of other DNA found in a cell.

Answers are now beginning to emerge. It turns out that many transcription factors include small projections called zinc fingers that are perfectly suited to DNA recognition. Our laboratory at the Medical Research Council in Cambridge, England, first identified a zinc finger in 1985—in a transcription factor obtained from a frog. Since then, more than 200 proteins, many of them transcription factors, have been shown to incorporate such zinc fingers. And several other transcription factors contain related structures, or motifs. Recently a number of laboratories, among them ours, have also begun to decipher just how zinc fingers and their relatives manage to select and grip their specific binding sites on DNA.

DANIELA RHODES and AARON KLUG both work at the Medical Research Council Laboratory of Molecular Biology in Cambridge, England. Rhodes, who holds a Ph.D. in biochemistry from the University of Cambridge, joined the council in 1969. She has been senior scientist since 1990. Klug, the 1982 winner of the Nobel Prize in Chemistry, began working at the Laboratory of Molecular Biology in 1962 and is now its director. The Nobel Prize recognized his development of electron microscopy techniques for determining the structure of complexes of biological molecules. It also honored his elucidation of the structure and assembly of protein-nucleic acid complexes in viruses and in chromosomes.

Of course, zinc fingers are not the only structures transcription factors exploit for interacting with DNA. Other important examples bear such names as helix-turn-helix motifs (discovered before zinc fingers, in 1981), homeodomains and leucine zippers [see "Molecular Zippers in Gene Regulation," by Steven Lanier McKnight; SCIENTIFIC AMERICAN, April 1991, and "Smart Genes," by Tim Beardsley, August 1991]. The zinc finger, however, is by far the most prevalent DNA-binding motif.

Ultimately, research into the problem of DNA recognition should advance inquiry into the larger question of how development unfolds in multicellular organisms. Although every cell in an embryo carries the same genes, some cells differentiate to become, say, neurons, whereas others become skin cells. Their fates vary because different combinations of genes are turned on in the cells as the embryo grows, leading to synthesis of the specialized proteins that give differentiated cells their distinctive properties. Knowledge of how transcription factors recognize their specific binding sites on DNA is central to an understanding of such selective gene activation.

We uncovered the existence of zinc fingers after becoming intrigued by results from the laboratories of Robert G. Roeder, then at Washington University, and Donald D. Brown of the Carnegie Institution of Washington in Baltimore. By 1980 Roeder and Brown and their associates had for the first time dissected the steps leading to transcription of a gene in an organism more advanced than bacteria. As part of that work, they demonstrated that in the frog *Xenopus laevis* a protein called transcription factor IIIA (TFIIIA) is one of at least three factors required to activate the gene that gives rise to 5S RNA. 5S RNA is a constituent of the ribosomes on which molecules of messenger RNA (the typical products of gene transcription) are translated into protein.

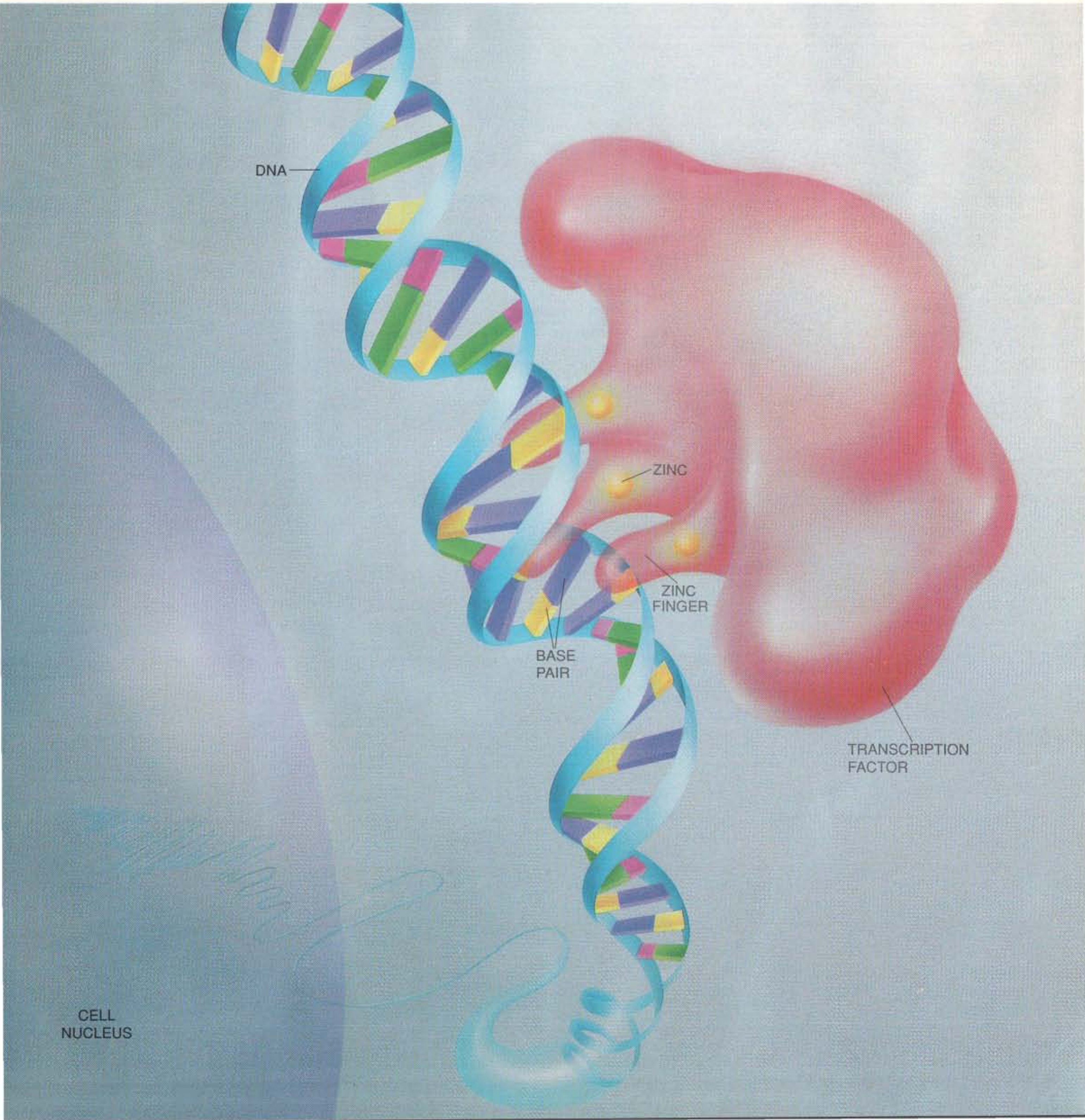
The investigators further found that TFIIIA binds to a relatively long patch of DNA, encompassing a particular sequence of about 45 base pairs, or "rungs" on the familiar DNA "ladder." (DNA is made up of two strands of nucleotides, which themselves consist of the sugar deoxyribose, a phosphate group and one of four distinguishing bases: adenine, cytosine, guanine or thymine. The two strands are attached to each other through their bases, so that adenine always pairs with thymine, and cytosine pairs with guanine.)

The length of the TFIIIA-docking site surprised us because TFIIIA is itself rather small. Transcription factors of the same size that had earlier been identified in bacteria attach themselves to much shorter tracts of DNA, on the order of 15 base pairs long. How, we asked, could this small TFIIIA molecule span such an extended stretch of DNA? Fortunately, the problem seemed tractable. Although transcription factors tend to be produced in scarce amounts, TFIIIA

THREE ZINC FINGERS (*protrusions*) extending from a transcription factor, or gene-regulating protein (*red*), have fastened themselves to the wide, major groove of a DNA molecule (*double helix*). Zinc fingers connect transcription factors to their target genes mainly by binding to specific sequences of DNA base pairs—the "rungs" in the twisted DNA "ladder." Zinc fingers are so named both because they can grasp DNA and because a zinc ion at the core (*yellow spheres*) plays a critical role in determining their structure.

is abundant in the ovaries of immature frogs. There it is stored as a complex with the 5S RNA it helps to generate. The abundance gave us confidence to think we could gather enough of the TFIIIA–5S RNA complex to isolate the protein. Having done that, we might learn something about the three-dimensional organization of the protein and about how it binds to its target site on the 5S RNA gene.

The plan was sound, but we soon encountered a difficulty—one that would prove fortunate because it set us directly on a path toward discovery of the zinc finger. In 1982 Jonathan Miller, then a research student in our laboratory, applied a known recovery technique to extract the TFIIIA–5S RNA complex from frog ovaries. He obtained disappointingly little of it. It turned out that the method he used was eliminating a metal needed to hold the complex together. After Miller modified the extraction procedure and procured a good supply of the complex, he showed that the lost metal was zinc. Each TFIIIA–5S RNA unit incorporated between seven and 11 zinc ions, an unusually large number. Other experiments led us further toward the zinc finger. When an enzyme called a protease chopped TFIIIA into successively smaller fragments, the fragments shrank by increments of about three kilodaltons (a measure of molecular weight). They ended up as three-kilodal-



ton units that resisted further attack, presumably because they were tightly folded. Collectively, these results suggested that TFIIIA was built almost entirely from a string of tandem three-kilodalton segments (representing about 30 amino acids per segment), each of which was folded around a zinc ion into a small, compact DNA-binding domain.

If we were right, the discovery would mean we had come across a novel kind of transcription factor. All others that had been studied in similar detail had been found to interact with DNA as dimers, or pairs, in which each protein in the dimer made contact with DNA through just one DNA-binding motif. Our findings implied that TFIIIA would stretch out along the double helix, touching it at several points instead of just one or two. Such multiple contacts would also explain how TFIIIA could interact with a very long segment of DNA.

As we were considering how to substantiate our model, Roeder's laboratory published the amino acid sequence of TFIIIA. In that sequence, we found support for our proposal: the first three quarters of the protein formed a continuous run of nine similar units of about 30 amino acids. Moreover, a pair of cysteine amino acids and a pair of histidine amino acids resided at virtually identical positions within each unit [*see bottom illustration in box at left*]. This last finding was consistent with the notion that each unit contained its own zinc ion, because zinc in proteins is generally found bound to four amino acids, often four cysteines or some combination of cysteines and histidines.

By 1985 these results led one of us (Klug) to propose formally that the invariant cysteines and histidines were used to fold each unit independently into a DNA-binding minidomain—later called a zinc finger because it was used to grip the DNA double helix. He suggested that the pair of cysteines near one end of the unit and the pair of histidines near the other end bound the same zinc atom, causing the intervening stretch of amino acids to loop out. Thus, in each 30-amino-acid unit, about 25 amino acids would fold into a structured domain (a finger); the remaining amino acids would serve as a linker between consecutive fingers [*see top illustration in box at left*].

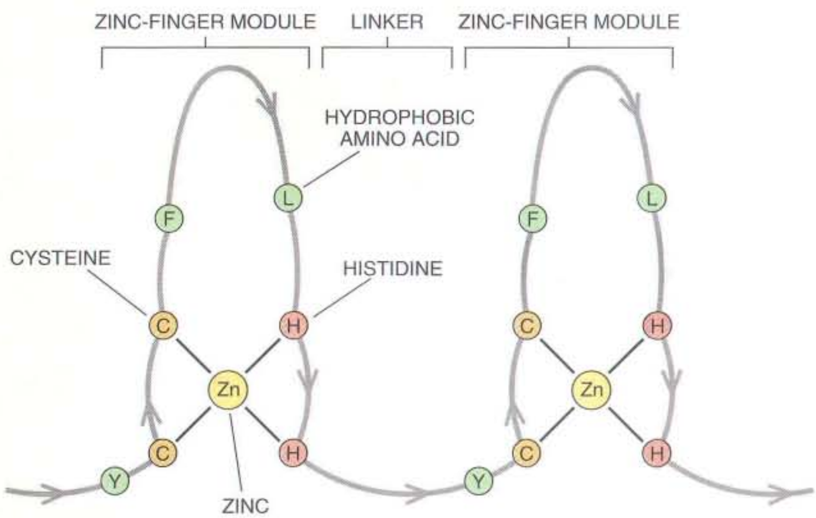
Shortly thereafter, measurements we made with Gregory P. Diakun of the Science and Engineering Research Council Daresbury Laboratory near Manchester, England, proved that each of the nine units did indeed contain a zinc ion bound to two cysteines and two histidines. TFIIIA was therefore fashioned almost entirely out of nine consecutive zinc fingers. All had the same basic architecture but were chemically distinct because of variations in the amino acids that did not participate in building the framework of the finger module.

But did the zinc fingers in fact contact DNA independently, as was predicted? To find out, Louise Fairall in our group, like investigators elsewhere, conducted what are called footprinting studies. A protein is allowed to attach to DNA. Then enzymes or other agents that attack DNA are applied. Any site that resists cleavage can be assumed to have been protected by the bound protein, indicating that the spared spot is a site of protein-DNA interaction. By 1986 the footprinting data confirmed that TFIIIA makes repeated contacts with DNA.

Hence, TFIIIA was the novelty we suspected it to be: it connected to a specific region on DNA by exploiting a string of independent DNA-binding modules. The economy of the modular arrangement was beautiful. Cells were already

How Zinc Fingers Were Discovered

One of the authors (Klug) deduced in 1985 that certain stretches of amino acids can fold independently around a zinc ion, forming modules that would come to be called zinc fingers (*bracketed regions at top*). The gray line represents a string of amino acids; the small, colored circles represent amino acids that Klug correctly thought might participate in the folding. A major clue (*bottom*) to the folding pattern came from inspection of the sequence of amino acids (*capital letters*) in the protein TFIIIA. The bulk of the protein can be arranged into nine successive sections, or sequence units (*numbered*), that exhibit important similarities: they include, at virtually identical positions, a pair of cysteine amino acids (*gold C's*), a pair of histidine amino acids (*red H's*) and (with the possible exception of section 7) three hydrophobic amino acids (*green letters*). (Asterisks mark unimportant breaks in the pattern.) These observations, added to biochemical findings, led to the proposal that the cysteine and histidine pairs in every module would combine with a single zinc ion (*large yellow spheres in top image*), causing the amino acids between those pairs to loop out as shown. At the same time, the three hydrophobic amino acids would somehow help stabilize the arrangement.



1 Y I C S F A D C G A A Y N K N W K L Q · A H L C · K H

2 T G E K · P F P C K E E G C E K G F T S L H H L T · R H S L · T H

3 T G E K · N F T C D S D G C D L R F T T K A N M K · K H F N R F H

4 N I K I C V Y V C H F E N C G K A F K K H N Q L K · V H Q F · S H

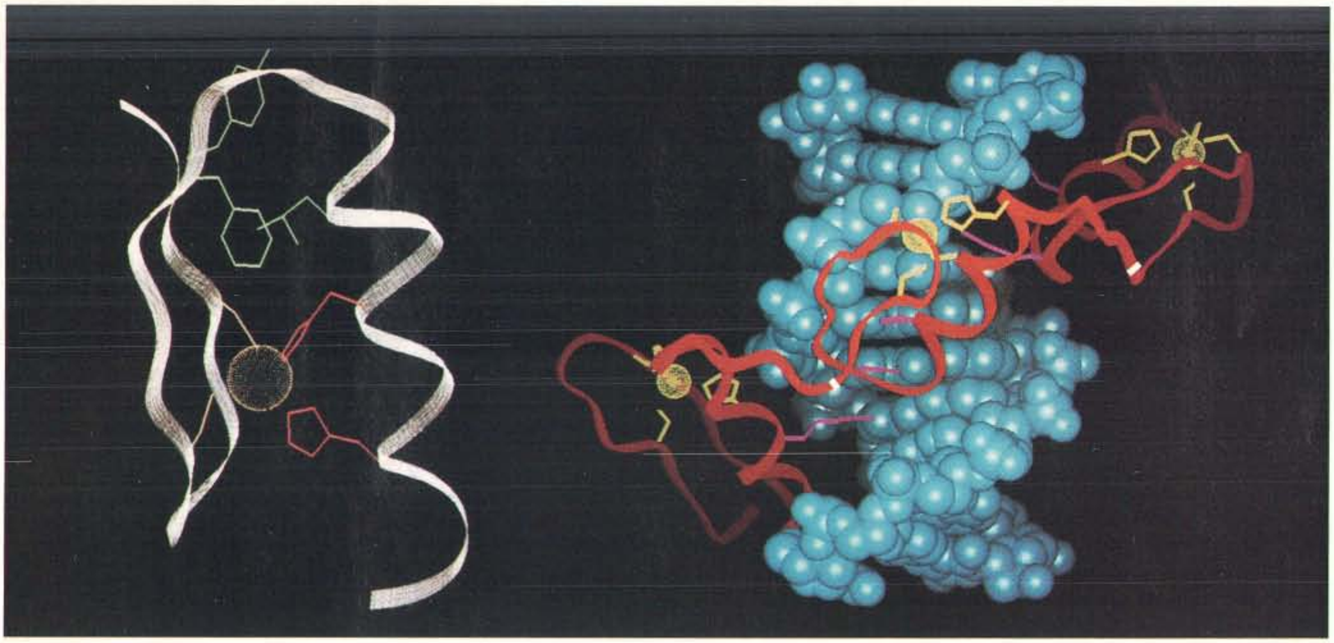
5 T Q Q L · P Y E C P H E G C D K R F S L P S R L K · R H E K · V H

6 A G · · · · Y P C K K D D S C S F V G K T W T L Y L K H V A E C H

7 Q D · · · L A V C · · D V C N R K F R H K D Y L R · D H Q K · T H

8 E K E R T V Y L C P R D G C D R S Y T T A F N L R · S H I Q S F H

9 E E Q R · P F V C E H A G C G K C F A M K K S L E · R H S V · V H



FINGER STRUCTURE (*left*) has been known in detail since the late 1980s. The ribbon represents the carbon-nitrogen backbone of the amino acid chain. The left half of the backbone folds back on itself to form a two-strand substructure known as a beta sheet (*V-shaped region*). The right half twists into an alpha helix (*spiral*). Binding of zinc (*yellow sphere*) by cysteines in the beta sheet (*yellow lines*) and histidines in the helix (*red lines*) draws the halves together near the base of the finger. It also brings hydrophobic amino acids (*green*) close to one another at the fingertip (*top of image*), where their mutual attraction helps to keep the structure intact. At the right, three tandem zinc fingers (*red ribbon trisected by white lines*) from the gene-regulating protein Zif268 have each made contact (*magenta*) with bases in the major groove of DNA (*blue*), collectively attaching to almost a full turn of the double helix. Five of the six base contacts are visible in this view. Yellow lines and rings represent the connections between zinc and the cysteines and histidines. The Zif268 image is based on an x-ray crystallographic analysis conducted at Johns Hopkins University by Nikola P. Pavletich and Carl O. Pabo.

known to build a large repertoire of on switches for genes by combining in various permutations a limited set of transcription factors. That is, one gene might be activated by a combination of proteins a, b and c, whereas another gene might make use of just a and b or of a, b and d. By such a strategy, organisms avoid having to produce a unique transcription factor for each of the enormous number of genes that are active in cells. The zinc-finger studies revealed that the combinatorial principle can also operate within a transcription factor. A cell can produce a vast collection of distinct transcription factors by varying the choice, order and number of independent DNA-binding modules in the proteins. The particular combination of zinc fingers in a transcription factor enables the factor to recognize a specific DNA sequence and no other.

The efficiency of the combinatorial approach led us to suggest that the zinc-finger motif might turn up in many proteins. But the extent of its occurrence in eukaryotes—organisms more advanced than bacteria—is astonishing. Peter F. R. Little of Imperial College, London, estimates that as much as 1 percent of the DNA in human cells specifies zinc fingers. In chromosome 19 the figure is as high as 8 percent. The zinc finger–containing proteins that have been identified so far carry from as few as two to as many as 37 tandem fingers.

To understand how a zinc finger recognizes a specific sequence of base pairs—which adopts a precise conformation—one needs to know the detailed three-dimensional structure of the finger module. Most proteins include local regions of "secondary" structure that fold together to yield the overall, three-dimensional shape of the protein. The most common secondary structures are the alpha helix (in which the backbone of the protein twists into a characteristic spiral) and the beta strand (in which the backbone is fully extended) [see "The Protein Folding Problem," by Frederic M. Richards; SCIENTIFIC AMERICAN, January 1991].

Jeremy M. Berg of Johns Hopkins University deciphered the important features of the three-dimensional architecture on theoretical grounds in 1988, but his model was not confirmed until 1989. Then Peter E. Wright and his colleagues at the Scripps Clinic and Research Foundation in La Jolla, Calif., determined the structure of a zinc finger from the *Xenopus* protein Xfin. They did it by applying nuclear magnetic resonance spectroscopy (NMR), a technique that can be used to solve the three-dimensional structure of small proteins in solution. Soon after, other laboratories and also our group identified the same design in other zinc-finger proteins.

As Berg predicted, the characteristic amino acid sequence of the zinc finger folds into a compact shape by forming two prominent substructures along the way. One part of the sequence (comprising, say, the left half of a vertical protrusion) adopts the shape of a small beta "sheet"; it is composed of two beta strands that form a sheet when the second strand folds back onto the first one [*see left illustration above*]. The other part of the sequence (the "right" half) twists into an alpha helix. The two cysteines reside at the bottom of the beta sheet, and the two histidines reside at the bottom of the helix. All four amino acids are joined through a zinc atom that essentially pins together the beta sheet and helix.

The NMR analysis also helped to clarify the role of a few additional amino acids. When originally examining the sequence of TFIIIA, we noted that the putative fingers each included a set of

three hydrophobic amino acids in virtually identical positions. (Hydrophobic substances often associate with one another in the interior of a protein in preference to water in the surroundings.) The invariance suggested that those amino acids had an important structural role. Although they are fairly far apart from one another in linear representations of the amino acid sequence, we thought they might somehow interact in three-dimensional space and thus assist in the folding of the minidomain. Consistent with the Berg model, the NMR results showed that when the zinc-finger module folds up, the hydrophobic amino acids do indeed come close enough to one another for their mutual attraction to come into play. They form a hydrophobic core that helps the module to maintain its shape.

In parallel with our efforts to understand the architecture of zinc fingers, we and others were also pondering a more general problem. Many experiments led to the conclusion that the zinc fingers in TFIIIA, which constitute the bulk of the protein, were solely responsible for the ability of the factor to recognize the promoter of the 5S RNA gene. But increasing numbers of proteins were being discovered in which only a few zinc fingers were embedded in a large protein. Could such short runs of zinc fingers direct these proteins to promoters, without assistance from other parts of the protein?

In our own efforts to answer these questions, we concentrated on a three-fingered yeast transcription factor called SWITCH 5 (SWI5). With our colleague Kyoshi Nagai, we isolated the region containing the fingers and exposed it to the promoter of the target gene for the protein. Sure enough, the isolated protein segment bound to the promoter avidly, implying that the zinc fingers are alone responsible for DNA binding. Interestingly, we found as well that at least two linked fingers had to be present for the SWI5 protein to attach with reasonable strength to its correct target site on DNA. By then applying NMR to the first two zinc finger motifs of SWI5, we and our colleagues David Neuhaus and Yukinobu Nakeseko confirmed that adjacent zinc fingers do not meld with each other; zinc fingers are truly independent "reading heads" joined by flexible linkers.

The precise points of contact between zinc fingers and DNA had yet to be identified, however. Nikola P. Pavletich and Carl O. Pabo, both then at Johns Hopkins, made the initial breakthrough in 1991. First, they obtained crystals of the complex formed by the DNA and the DNA-binding domain of a transcription factor called Zif268. By then carrying out an x-ray crystallographic analysis, they were able to determine the detailed structure of the complex. Zif268, which in common with SWI5 includes a run of three zinc fingers, participates in the early development of mice.

The x-ray analyses revealed that the Zif268 zinc-finger region curls around almost one turn of the DNA helix (more or less tracing the letter "C"), fitting itself into the major groove. (The major groove is the wider of two parallel gullies that spiral around the long axis of the DNA double helix, much as red-and-white ribbons of color encircle old-fashioned barber poles.) The fingers make contacts with successive, three-base-pair sites on the DNA, and they approach the DNA in much the same orientation. That is, the alpha helix of each finger points into the major groove, abutting one of its walls.

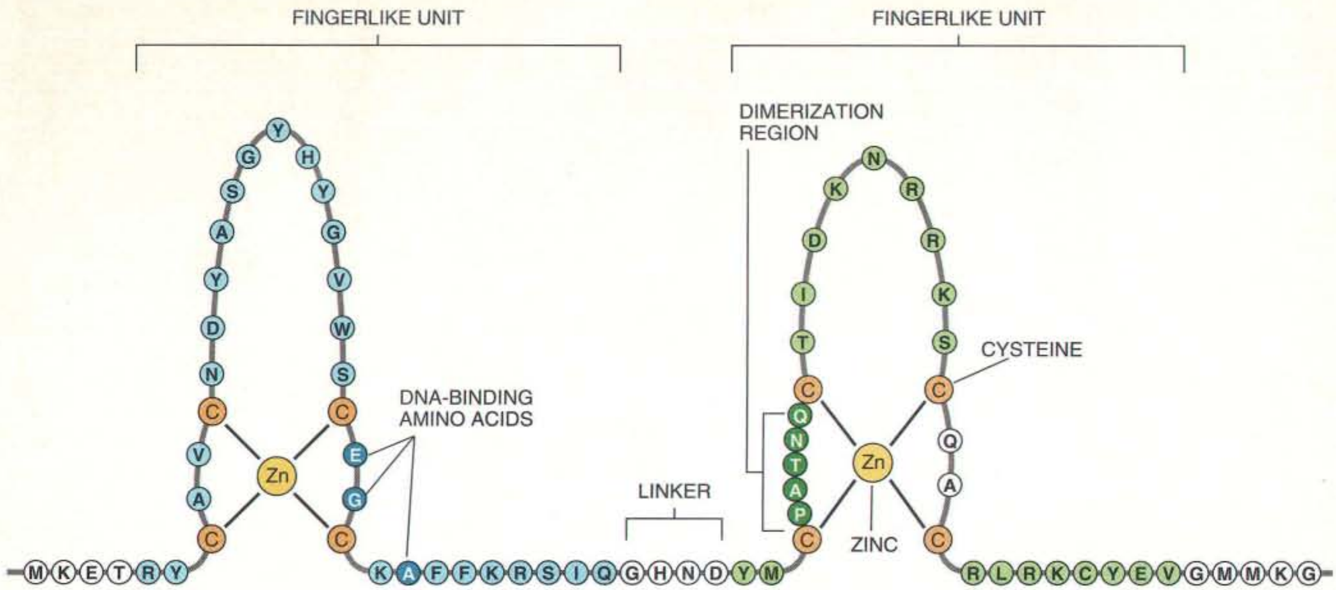
More specifically, the first and third fingers of Zif268 bind to DNA identically: an amino acid in the first turn of the alpha helix contacts the first base pair of the corresponding binding site on DNA, and an amino acid in the third turn of the helix contacts the third base pair of that same DNA site. The second finger also makes two contacts through the alpha helix, but this time amino acids on the first and second turn contact the first and second base pairs of the corresponding binding site on DNA. (In each instance, one amino acid contacts one DNA base in a pair.) In addition, both the alpha helix and the beta sheet in the fingers bind to phosphate groups in the chains of sugar and phosphate that make up the "sides" of the DNA ladder. These added links help to stabilize the attachment of zinc fingers to DNA.

So far no other complexes of zinc fingers and DNA have been solved by x-ray crystallography. Nevertheless, Grant H. Jacobs in our laboratory has good evidence that many zinc fingers bind to DNA in much the same way as Zif268 does. Jacobs has compared the amino acid sequences of more than 1,000 zinc-finger motifs. He finds that the amino acids in three positions are particularly variable. These highly variable positions are precisely those that are used to make contacts in the Zif268 complex, namely, those falling on the first, second and third turns of the alpha helix. Such similarity raises the exciting possibility that zinc-finger modules might one day be designed at will to recognize selected DNA sequences—a feat that could be important both for the study of gene regulation and for medicine.

Of course, there are limits to how much one can extrapolate from the Zif268 model and from statistical analyses. Proteins with many zinc fingers would be expected to interact with DNA somewhat differently. For instance, if the Zif268 pattern of binding applied to TFIIIA, this protein, with its nine fingers, would wind around the DNA for three turns, like thread on a spool. This extensive wrapping could well hamper the factor from coming off the DNA when detachment became necessary. Indeed, footprinting data obtained by us and others suggest that TFIIIA does not twist continuously around the DNA. The first three fingers of TFIIIA almost certainly clasp onto a single turn of the DNA, and it is very likely that the last three fingers do the same. But the bulk of the protein lies on just one face of the double helix; hence, it crosses the narrow, minor groove at least twice. The varied DNA-binding patterns of separate regions of TFIIIA probably reflect the fact that the amino acid sequences of the TFIIIA fingers differ more from one another than do those of the fingers in Zif268-type proteins.

From an evolutionary standpoint, there is good reason to think that multifingered DNA-binding domains arose by duplication of some ancestral gene that specified a small protein of 30 or so amino acids. We think, too, that the 30-amino-acid chain may have been among the earliest of proteins to evolve. Such a protein would, after all, have been simple to produce. Once synthesized, it would easily and safely pick up zinc (which is a relatively inert metal) from its surroundings and would then fold without assistance into a stable conformation. So folded, it would acquire the ability to bind to DNA or RNA. Such attributes almost certainly help to explain why zinc fingers are now prevalent throughout the animal and plant kingdoms. Any species acquiring the genetic blueprint for a particular, autonomously folding zinc finger would instantly acquire the ability to bind to a new stretch of DNA. That property, in turn, could give rise to new cellular functions, such as the ability to transcribe some previously silent gene and thus to produce a novel enzyme or other valuable protein.

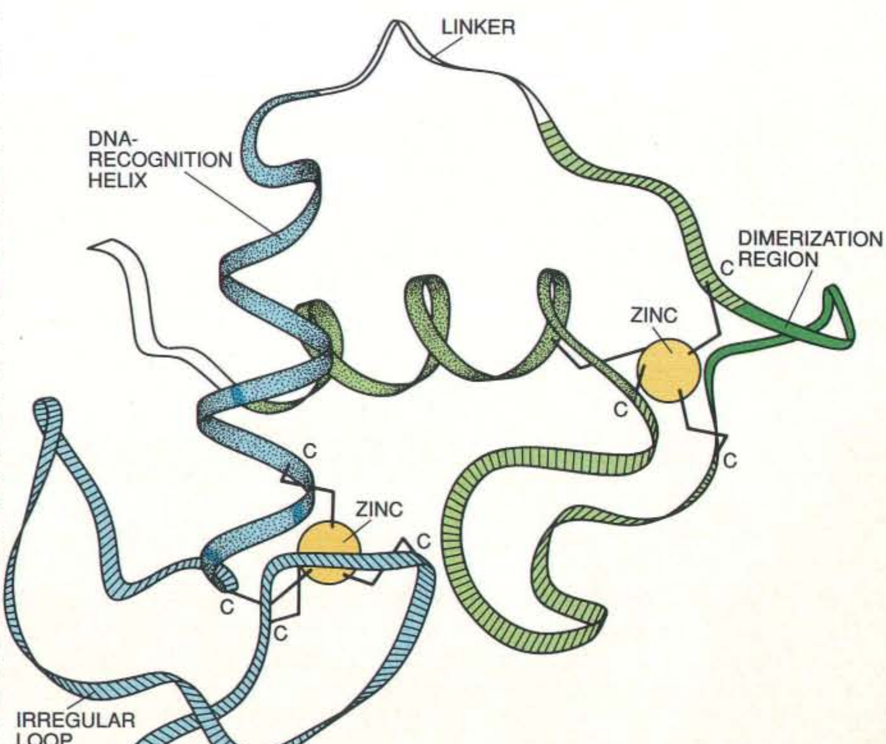
As we and others were gaining insight into the structure and function of classic zinc fingers, tantalizing findings began to suggest that the motif we had initially discovered in TFIIIA was not the only zinc-centered structure devoted to DNA recognition.



DNA-BINDING DOMAIN of the estrogen receptor (a transcription factor that must be bound by estrogen to act on a gene) is built from an amino acid sequence (*capital letters*) that has here been divided into two zinc-finger-like units (*blue and green regions*). Such diagrams initially led investigators to assume that the two units, like classical zinc fingers, each recognize separate base sequences in the DNA. In fact, only the three amino acids colored dark blue are thought to interact with DNA bases, which means the first fingerlike unit makes the main contact with DNA. The second unit serves a different function than the first: it carries five amino acids (*dark green*) that enable one receptor molecule to combine, or dimerize, with a second receptor molecule. Such pairing is required if estrogen receptors are to attach securely to DNA.

Three-Dimensional View of the DNA-Binding Domain of the Estrogen Receptor

The detailed structure of the DNA-binding domain of the estrogen receptor was determined in 1990 by one of the authors (Rhodes) and her collaborators John W. R. Schwabe and David Neuhaus. This work and biochemical data enabled them to pinpoint the parts of the three-dimensional structure that perform the critical tasks of recognizing DNA and pairing with other receptor molecules. The two units in the domain (*light blue and light green*) adopt similar conformations. An irregularly structured loop (*hatched*) that includes two cysteines (*C's*) is followed by an alpha helix (*stippled spiral*) carrying the third and fourth cysteines. Binding of zinc (*yellow*) by the cysteines ties the terminal segments of the irregular region to the base of the helix. So folded, the two units mesh through their helices. The amino acids responsible for recognizing specific bases (*dark blue*) fall on the helix in the first unit; those responsible for forming a dimer (*dark green*) reside in the irregular loop of the second unit.



By 1987 investigators had elucidated the amino acid sequences of several members of a large family of transcription factors known as nuclear hormone receptors. Such factors must be bound by a particular steroid or thyroid hormone or vitamin before they can activate a gene. In examining the newly determined sequences, workers saw that every one of them bore a conserved, or highly similar, domain of about 80 amino acids. This domain consistently included two, and always two, units whose amino acid sequence was reminiscent of the zinc finger. As was true of zinc fingers, each unit, or motif, contained two pairs of potential zinc-binding amino acids; here, however, the zinc binders were exclusively cysteines instead of cysteines and histidines. These resemblances of the sequences to TFIIIA zinc-finger motifs implied that the cysteine-rich, 80-amino-acid segment of the factors was the DNA-binding domain.

Pierre Chambon and Stephen Green of INSERM in Strasbourg confirmed that assumption in the late 1980s. Soon after, Paul B. Sigler, then at the University of Chicago, and Keith R. Yamamoto of the University of California at San Francisco and their associates established that each of the two segments of the DNA-binding domain incorporates a zinc atom. Naturally, we and others expected that, as is true of TFIIIA-type zinc fingers, the configurations of the two motifs would resemble each other, and the motifs would form independent DNA-binding modules.

The assumption turned out to be partly wrong. Structural analyses would eventually show that the two units do fold similarly. But, before that, some striking biochemical work would demonstrate that the units do not function as independent DNA-reading heads. By substituting one amino acid for another and examining the effect on DNA binding, Chambon, Ronald M. Evans of the Salk Institute for Biological Studies in San Diego and Gordon M. Ringold, formerly of Stanford University, and their associates found that the first motif serves as the primary DNA-recognition unit. At about the same time, Evans and his co-worker Kazuhiko Umesono, again applying the substitution method, uncovered at least one function of the second motif. To understand that function, one must first know something general about how steroid receptors interact with DNA.

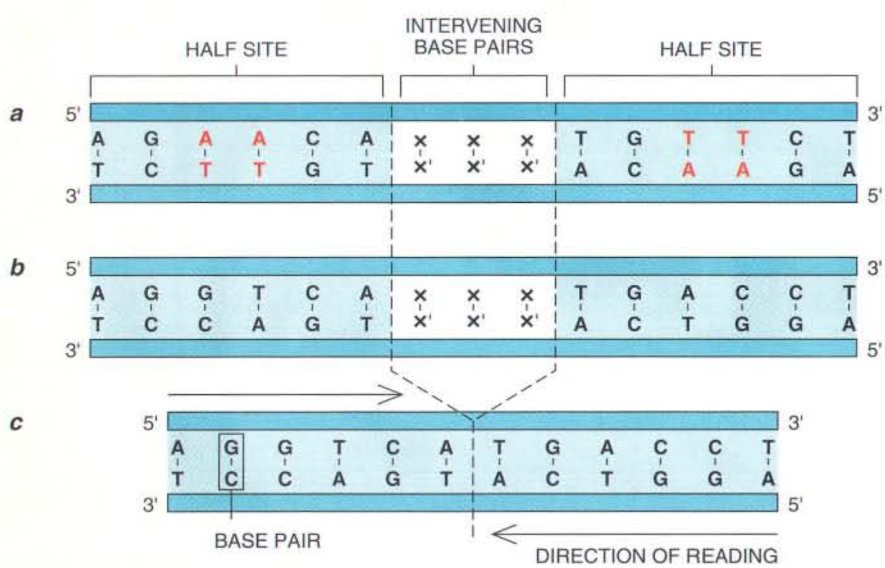
Such receptors bind to DNA as pairs, or dimers, of identical molecules. Each protein in a pair recognizes half of a two-part binding site that is known as a palindrome, because the halves are identical if read in opposite directions—that is, along opposite strands of the DNA [*see illustration below*]. The base sequence of the half site recognized by one type of transcription factor (such as the estrogen receptor) can exactly match that recognized by another factor (such as the thyroid receptor). In that case, the only difference between the two binding sites is the number of base pairs separating the half sites in each palindrome.

Consequently, for a transcription factor to find its corresponding docking site on DNA, the protein must contain regions dedicated to picking out a specific half-site base sequence and also to measuring the distance between half sites. Evans and Umesono found that a part of the second motif is responsible for measuring such spacing.

Despite this progress, one could not fit these pieces of information together to explain how sequence-specific recognition took place. That explanation could come only from knowing the three-dimensional configuration of the DNA-binding domains of receptor proteins and thus seeing where on the structure the functionally important amino acids would lie. In 1990, by applying NMR, Robert Kaptein and his colleagues at the University of Utrecht solved the structure of the DNA-binding domain of the rat glucocorticoid (cortisone) receptor. Shortly thereafter, John W. R. Schwabe and Neuhaus in our laboratory and one of us (Rhodes) solved that of the human estrogen receptor.

As could be deduced from their similar amino acid compositions, the DNA-binding domains of the glucocorticoid and estrogen receptors were found to adopt much the same structure. Each of the two zinc finger–containing motifs within the domain consists of two parts: an irregularly looped string of amino acids (instead of the beta sheet in classic zinc fingers), followed by an alpha helix. The loop carries two of the zinc-binding sites, and the other two reside at the beginning of the helix that follows. Yet instead of remaining separate, as standard zinc fingers would, the two motifs merge into a single structural unit. In this arrangement, the helices cross perpendicularly at their midpoints, a configuration that is created by the mutual attraction of invariant and relatively invariant hydrophobic amino acids.

With the three-dimensional structure of the DNA-binding domain known, we proceeded to map onto it the locations of amino acids that had earlier been shown to be critical for DNA recognition. Groups led by Chambon, Evans and Ringold had identified three amino acids in the first fingerlike motif that were responsible for recognizing the base sequence of a half site. Those amino acids turn out to reside on one face of the helix in the motif, leading

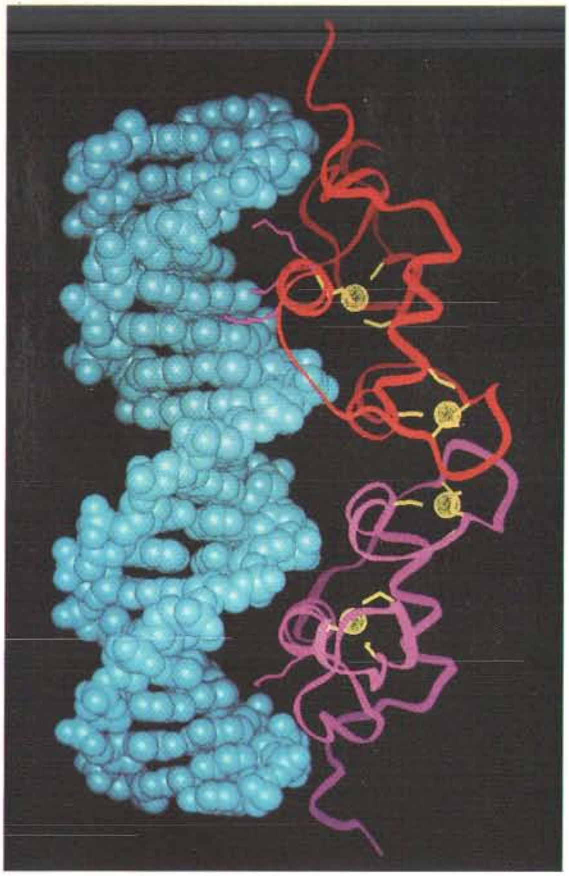


DOCKING SITES ON DNA, or response elements, that are recognized by the glucocorticoid (*a*), estrogen (*b*) and thyroid (*c*) receptors include two half sites (*shaded regions*). The half sites in any element are alike, if their base pairs (*linked letters*) are read along opposite strands of the DNA (*arrows in c*), in the 5′ to 3′ direction. Receptors bind to the response elements as dimers, or pairs—one molecule to each half site. To bind successfully, they must be able to distinguish both the base sequence of, and the spacing between, half sites. The differences between response elements can be subtle: element *b* differs from *a* by only two base pairs (*red letters*), and it differs from *c* only in the number of bases separating the half sites.

us to call that substructure the DNA-recognition helix. This information also told us something more about the function of the second motif: by crossing the DNA-recognition helix of the first motif, the helix of the second motif serves as a backing strut to hold the recognition helix in place. The separation of function between the two motifs suggests that the second finger arose from duplication of the first but that once the second unit appeared, it was pressed into service for new tasks.

Mapping in three dimensions told us as well how the second motif performs the vital role of discriminating the spacing between half sites on DNA. Evans and Umesono had established that the amino acids responsible for such discrimination lie between the first two cysteines of the second motif. In the three-dimensional configuration, these amino acids map to the loop preceding the helix, where they would be available to link one molecule to its partner. Computer modeling of the interaction between DNA and the DNA-binding regions of the glucocorticoid and estrogen receptors then enabled Kaptein's and our group, respectively, to see that the pairing of proteins through the predicted connection would orient the dimer properly. The two recognition helices on the dimer would be arranged so that the spacing between them would match the spacing between the appropriate half sites in the DNA.

Sigler and his colleague Ben F. Luisi, working together at Yale University, in collaboration with Yamamoto and Leonard P. Freedman of the University of California at San Francisco have since confirmed this picture by x-ray crystallography. They have also learned that each protein in the dimer makes several contacts with the phosphates on either side of the major groove. These contacts position the DNA-recognition helix so that it can reach deep into the major groove to form bonds with base pairs in the half site. Overall, then, studies of the nuclear hormone receptor class of zinc-finger motifs indicate that, despite some structural similarity to TFIIIA-type zinc fingers, these motifs function more like the DNA-recognition motifs of other transcription factors, such as the helix-turn-helix and leucine zipper. That is, by folding together instead of remaining distinct, the motifs help nuclear hormone receptors to form the dimers that enable such factors to recognize their specific binding sites on DNA.



ATTACHMENT TO DNA has been accomplished by the DNA-binding domains (*red and magenta ribbons*) of a pair of glucocorticoid receptors. One alpha helix (*spiral*) from each domain makes contact (*thin magenta lines*) with bases on a single face of the DNA double helix. The image is based on an x-ray crystallographic analysis carried out by Ben F. Luisi and Paul B. Sigler and their colleagues at Yale University.

When knowledge of the structure of a molecule reveals something about the way in which it works, that information may also offer insight into disease. In the case of zinc fingers, researchers have learned that a renal cancer called Wilm's tumor arises from a genetic mutation that interferes with the proper binding to DNA of the zinc-finger region in a particular protein. Moreover, some of the symptoms that can follow from insufficient intake of zinc in the diet—such as delayed sexual development—can now be attributed to the inability of estrogen and androgen receptors to fold properly in the absence of zinc.

Clearly, the two classes of zinc fingers we have discussed vary profoundly in both their structure and the way in which they interact with DNA. We have no doubt that still more variety will be discovered in the extended family of zinc-finger proteins. Nature continues to surprise and amuse us with the ingenuity of the designs it has evolved to enable proteins to recognize specific base sequences in DNA. For instance, an increasing number of amino acid sequences include what seem to be zinc-binding motifs, although the spacing between the pairs of cysteines or histidines, or the numbers of pairs, differs from that in the standard zinc finger. One unusual example is the yeast protein GAL4; it bears six cysteines that fold around two zinc atoms.

We also expect to find that some zinc fingers or their cousins are involved in activities other than transcription, such as transporting, processing or otherwise acting on DNA or even RNA; recall, for instance, that TFIIIA binds to RNA as well as to DNA. We still have much to learn.

FURTHER READING

REPETITIVE ZINC-BINDING DOMAINS IN THE PROTEIN TRANSCRIPTION FACTOR IIIA FROM *XENOPUS* OOCYTES. J. Miller, A. D. McLachlan and A. Klug in *EMBO Journal*, Vol. 4, No. 6, pages 1609–1614; 1985.

ZINC FINGERS: A NOVEL PROTEIN MOTIF FOR NUCLEIC ACID RECOGNITION. Aaron Klug and Daniela Rhodes in *Trends in Biochemical Sciences*, Vol. 12, No. 12, pages 464–469; December 1987.

ZINC FINGER—DNA RECOGNITION: CRYSTAL STRUCTURE OF A ZIF268-DNA COMPLEX AT 2.1Å. Nikola P. Pavletich and Carl O. Pabo in *Science*, Vol. 252, pages 809–817; May 10, 1991.

BEYOND ZINC FINGERS: STEROID HORMONE RECEPTORS HAVE A NOVEL STRUCTURAL MOTIF FOR DNA RECOGNITION. John W. R. Schwabe and Daniela Rhodes in *Trends in Biochemical Sciences*, Vol. 16, No. 8, pages 291–296; August 1991.

CRYSTALLOGRAPHIC ANALYSIS OF THE INTERACTION OF THE GLUCOCORTICOID RECEPTOR WITH DNA. B. F. Luisi, W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto and P. B. Sigler in *Nature*, Vol. 352, No. 6335, pages 497–505; August 8, 1991.

A STRUCTURAL TAXONOMY OF DNA-BINDING DOMAINS. Stephen C. Harrison in *Nature*, Vol. 353, No. 6346, pages 715–719; October 24, 1991.

How Should Chemists Think?

Chemists can create natural molecules by unnatural means. Or they can make beautiful structures never seen before. Which should be their grail?

by Roald Hoffmann

The Vatican holds a fresco by Raphael entitled *The School of Athens.* Plato and Aristotle stride toward us. Plato's hand points to the heavens, Aristotle's outward, along the plane of the earth. The message is consistent with their philosophies—whereas Plato had a geometric prototheory of the chemistry of matter, Aristotle described in reliable detail how Tyrian purple (now known to be a precursor of indigo) was extracted from rock murex snails. Plato searched for the ideal; Aristotle looked to nature.

Remarkably, modern chemistry faces the quandary that Raphael's fresco epitomized. Should it follow the hand sign of Aristotle or that of Plato? Is nature as fertile a source for new materials as some assert it to be? Can we, for example, hope to make better composites by mimicking the microstructure of a feather or of a strand of spider's silk? Are chemists better advised to seek their inspiration in ideal mathematical forms, in icosahedra and in soccer balls? Or should we hazard chance?

To some, the division between natural and unnatural is arbitrary; they would argue that man and woman are patently natural, and so are all their transformations. Such a view is understandable and has a venerable history, but it does away with a distinction that troubles ordinary and thoughtful people. So I will not adopt it and instead will distinguish between the actions, mostly intended, of human beings and those of animals, plants and the inanimate world around us. A sunset is natural; a sulfuric acid factory is not. The 1.3 billion head of cattle in this world pose an interesting problem for any definition. Most of them are both natural and unnatural—the product of breeding controlled by humans.

The molecules that exist naturally on the earth emerged over billions of years as rocks cooled, oceans formed, gases escaped and life evolved. The number of natural molecules is immense; perhaps a few hundred thousand have been separated, purified and identified. The vast majority of the compounds that fit into the unnatural category were created during the past three centuries. Chemists have added some 15 million well-characterized molecules to nature's bounty.

To every thing of this world, be it living or not, there is structure. Deep down are molecules, persistent groupings of atoms associated with other atoms. There is water in the distilled form in the laboratory, in slightly dirty and acid snow, in the waters associated with our protein molecules. All are H_2O. When chemistry was groping for understanding, there was a reasonable reluctance to merge the animate and inanimate worlds. Friedrich Wöhler convinced many people that the worlds were not separate by synthesizing, in 1828, organic urea from inorganic silver cyanate and ammonium chloride.

How are molecules made in nature—penicillin in a mold or a precursor of indigo in a rock murex snail? How are they made in glass-glittery laboratories—those acres of food wrap, those billion pills of aspirin? By a common process—synthesis.

Chemistry is the science of molecules and their transformations. Be it natural or human-steered, the outcome of transformation, $A \rightarrow B$, is a new substance. Chemical synthesis, the making of the new, is patently a creative act. It is as much an affirmation of humanity as a new poem by A. R. Ammons or the construction of democracy in Russia. Yet creation is always risky. A new sedative may be effective, but it also may induce fetal malformation. A Heberto Padilla poem may be "counterrevolutionary" to a Cuban apparatchik. Some people in Russia still don't like democracy.

Wöhler mixed together two substances, heated them and obtained an unexpected result. Much has happened since 1828. To convey what the making of molecules is like today and to relate how the natural intermingles with the unnatural in this creative activity, let me tell you about the synthesis of two substances: Primaxin and the ferric wheel.

Primaxin is one of the most effective antibiotics on the market, a prime money-maker for Merck & Co. The pharmaceutical is not a single molecule but a designed mixture of two compounds, imipenem and cilastatin [*see box on page 42*]. These are their "trivial" names. The "systematic" names are a bit longer; for instance, imipenem is

[5R-[5α, 6α(R*)]]-6-(1-hydroxyethyl)-3-[[2-[(iminomethyl)amino]ethyl]thio]-7-oxo-1-azabicyclo[3.2.0]hept-2-ene-2-carboxylic acid.

Primaxin was created by a bit of unnatural tinkering, emulating the natural tinkering of evolution. Imipenem by itself is a fine antibiotic. But it is degraded rapidly in the kidney by an enzyme. This would give the drug limited use for urinary tract infections. The Merck chemists found in their sample collection a promising compound, synthesized in the 1940s, that inhibited that ornery enzyme. Modified for greater activity, this became cilastatin. It was obvious to try the combination of the antibiotic and the enzyme inhibitor, and the mix worked.

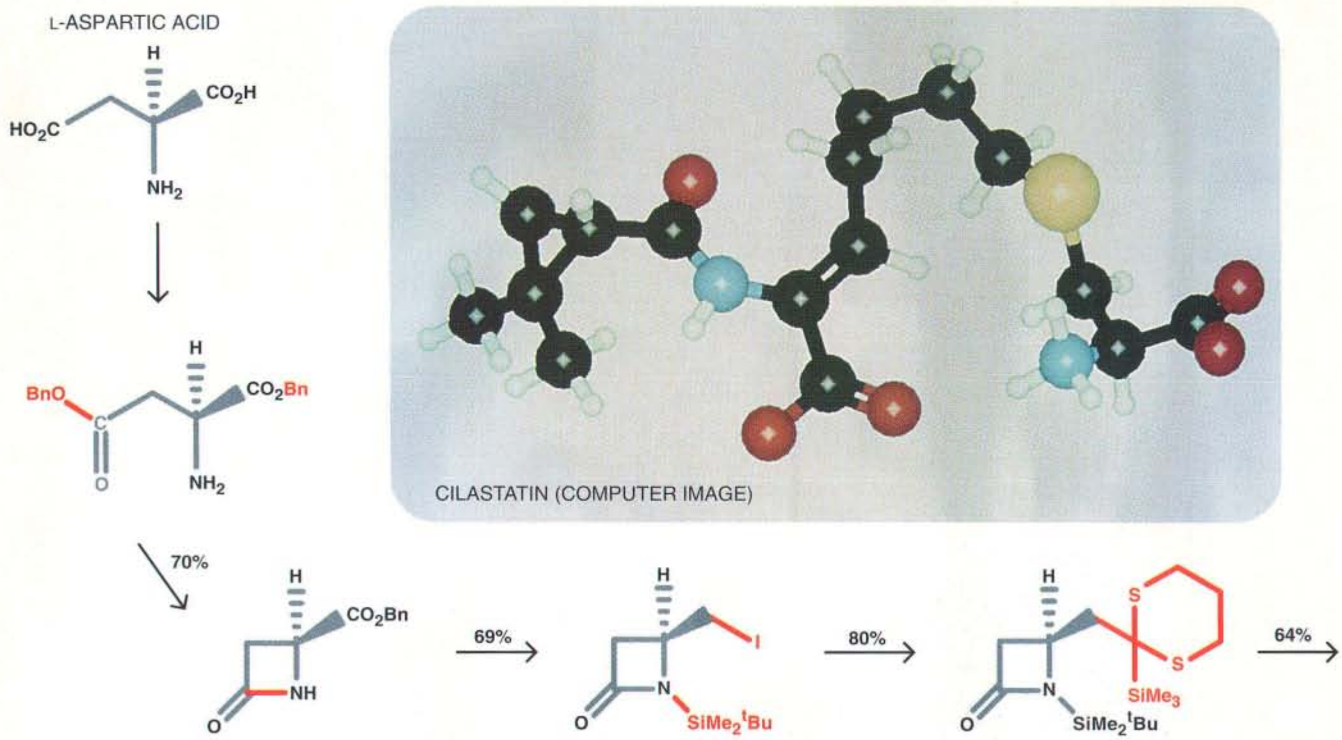
Imipenem derives from a natural

ROALD HOFFMANN shared the 1981 Nobel Prize in Chemistry with Kenichi Fukui. He was born in 1937 in Zloczow, Poland. Having survived the war, he came to the U.S. in 1949. He studied chemistry at Columbia University and received his Ph.D. from Harvard University. In 1965 he joined the faculty at Cornell University and is now the John A. Newman Professor of Physical Science. Hoffmann describes his contribution to science as "applied theoretical chemistry"—a particular blend of computations stimulated by experiment and the construction of generalized models. He writes essays and has published two poetry collections, *The Metamict State* and *Gaps and Verges.*

PLATO AND ARISTOTLE in a detail of Raphael's fresco *The School of Athens* are depicted in a way that symbolizes their approach to knowledge. Aristotle gestures toward the earth; Plato points his finger to the heavens. Aristotle looked to nature for answers; Plato searched for the ideal. Should chemists follow the hand sign of Aristotle or that of Plato?

The Making of an Antibiotic

The antibiotic Primaxin is a mixture of two compounds known as imipenem (*ball-and-stick model at right*) and cilastatin (*model at left*). Imipenem is a slightly modified form of thienamycin, which is produced naturally by a mold. Chemists developed a procedure (summarized below) that produces thienamycin more efficiently than any known natural process. The stick figures shown are the chemist's typical notation; not all atoms are identified. Those vertices that do not have atomic labels represent carbon atoms. Most of the hydrogen atoms have been left out. It is possible to de-



CILASTATIN (COMPUTER IMAGE)

product; cilastatin does not. Both are made synthetically in the commercial process. I will return to this after tracing further the history of one of the components.

Imipenem was developed in the 1970s by a team of Merck chemists led by Burton G. Christensen. It is a slightly modified form of another antibiotic, thienamycin. That, in turn, was discovered while screening soil samples from New Jersey. It is produced by a mold, *Streptomyces cattleya,* so named because its lavender color resembles that of the cattleya orchid. The mold is a veritable drug factory, producing thienamycin and several other varieties of antibiotics.

Unfortunately, thienamycin was not chemically stable at high concentrations. And, to quote one of the Merck crew, "The lovely orchid-colored organism was too stingy." The usual fermentation processes, perfected by the pharmaceutical industry over the past 50 years, did not produce enough of the molecule. So the workers decided to produce greater quantities of thienamycin in the laboratory.

The production of thienamycin required 21 major steps, each involving several physical operations: dissolution, heating, filtration, crystallization. Between the starting material—a common amino acid, L-aspartic acid—and the desired product—thienamycin—20 other molecules were isolated and purified. Of these, only eight are shown in the condensed "reaction scheme" above.

The first impression that one gets is of complexity. That intricacy is essential, a laboratory counterpoint to the biochemical complexity of bacteria and us. We would like there to be "magic bullets" of abiding simplicity. The real world is complicated and beautiful. We had better come to terms with that richness.

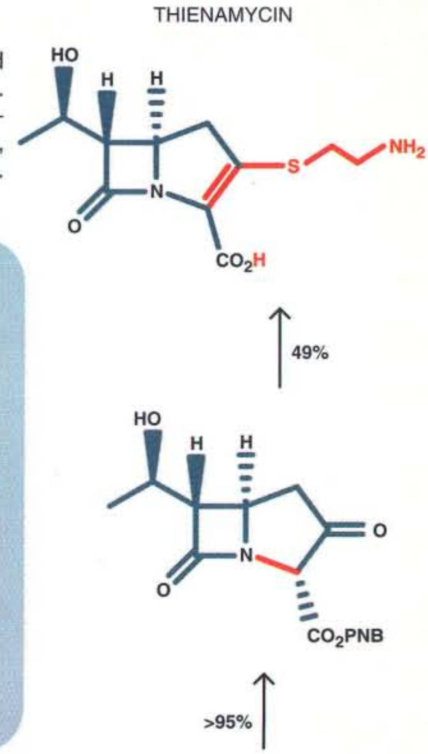
To get a feeling for the sweat, if not the blood and tears, of the process, we need to turn to the experimental section of the paper reporting the synthesis. Here is an excerpt of that experimental protocol, describing a critical, inventive step in the synthesis—the transformation from compound 8 to 9:

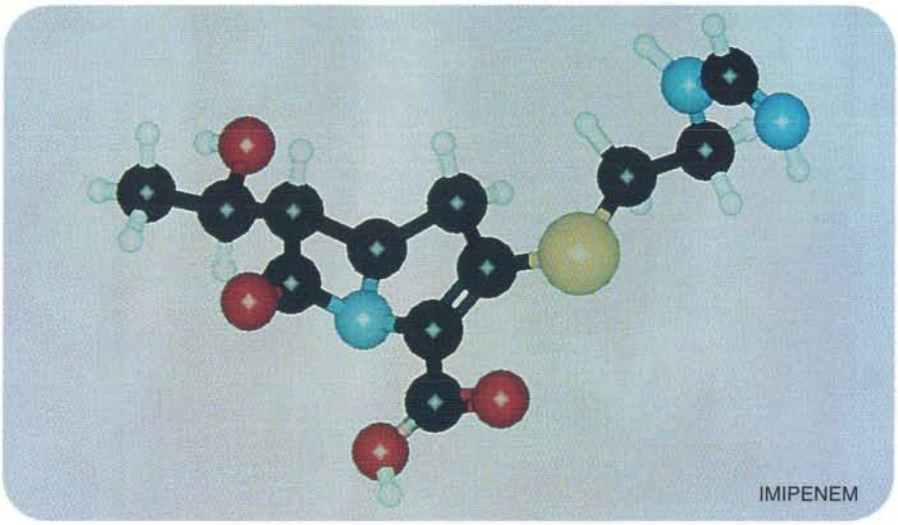
> A suspension of diazo keto ester 8 (3.98g, 10.58 mmol) and rhodium(II) acetate dimer (0.04 g, 0.09 mmol) in anhydrous toluene (250 mL) was thoroughly purged with nitrogen, and then heated with stirring in an oil bath maintained at 80°C. After heating for two hours, the reaction mixture was removed from the bath and filtered while warm through a pad of anhydrous magnesium sulfate. The filtrate was evaporated under vacuum to afford the bicyclic keto ester 9 (3.27 g, 89%) as an off-white solid....

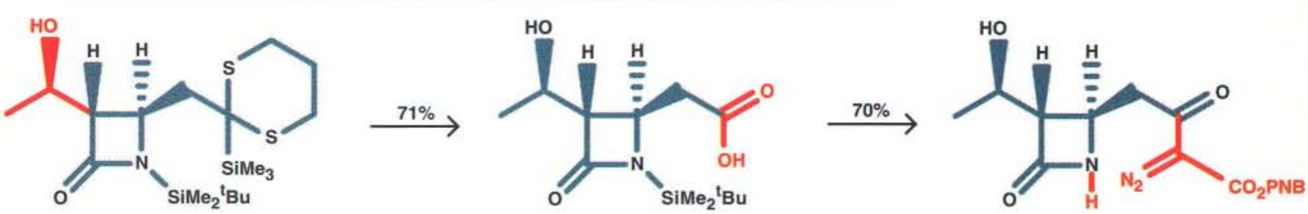
You can be sure that this jargon-laden account of an experimental procedure is a sanitized, too linear narrative; it is the way things were at the end: neat, optimized. Not the way it first happened. Putting that aside, you feel work, a sequence of operations that take time and effort. Sometimes, just as in our romantic notions of words springing from the brow of inspired poets, we forget the sheer labor of creation. Even the Creator rested on the seventh day.

You might be interested to see the way these experimental procedures change when the very same process is

duce the location of the missing hydrogen atoms because every carbon atom should form four bonds. An arrow represents each chemical transformation in the process. The percent figure near each arrow is the experimental yield. The symbols highlighted in red indicate changes in the structure. In the models of cilastatin and imipenem, black is carbon, blue is nitrogen, yellow is sulfur, red is oxygen and white is hydrogen. Wedges indicate details of geometry, atoms above or below the plane.







scaled up. You can't make hundreds of millions of dollars' worth of thienamycin the same way you make a few grams in the laboratory. Here is the description of the industrial synthesis, for the very same step:

> The solids containing 200 kg of 8 are dropped into 476 gallons of $MeCl_2$ in tank TA-1432. Meanwhile, the reactor ST-1510 is cleaned out by a 200-gallon $MeCl_2$ boilout. The slurry is transferred to ST-1510, followed by a 50-gallon $MeCl_2$ line flush. An additional 400 gallons of dry $MeCl_2$ are added to ST-1510, and hot water (65°C) is applied to the jackets to concentrate the batch to 545 gallons where the slurry KF (Karl Fischer) is approximately 0.5 g/l H_2O. Distillates are condensed and collected in another tank.

Making veal stroganoff for a thousand people is not the same as cooking at home for four.

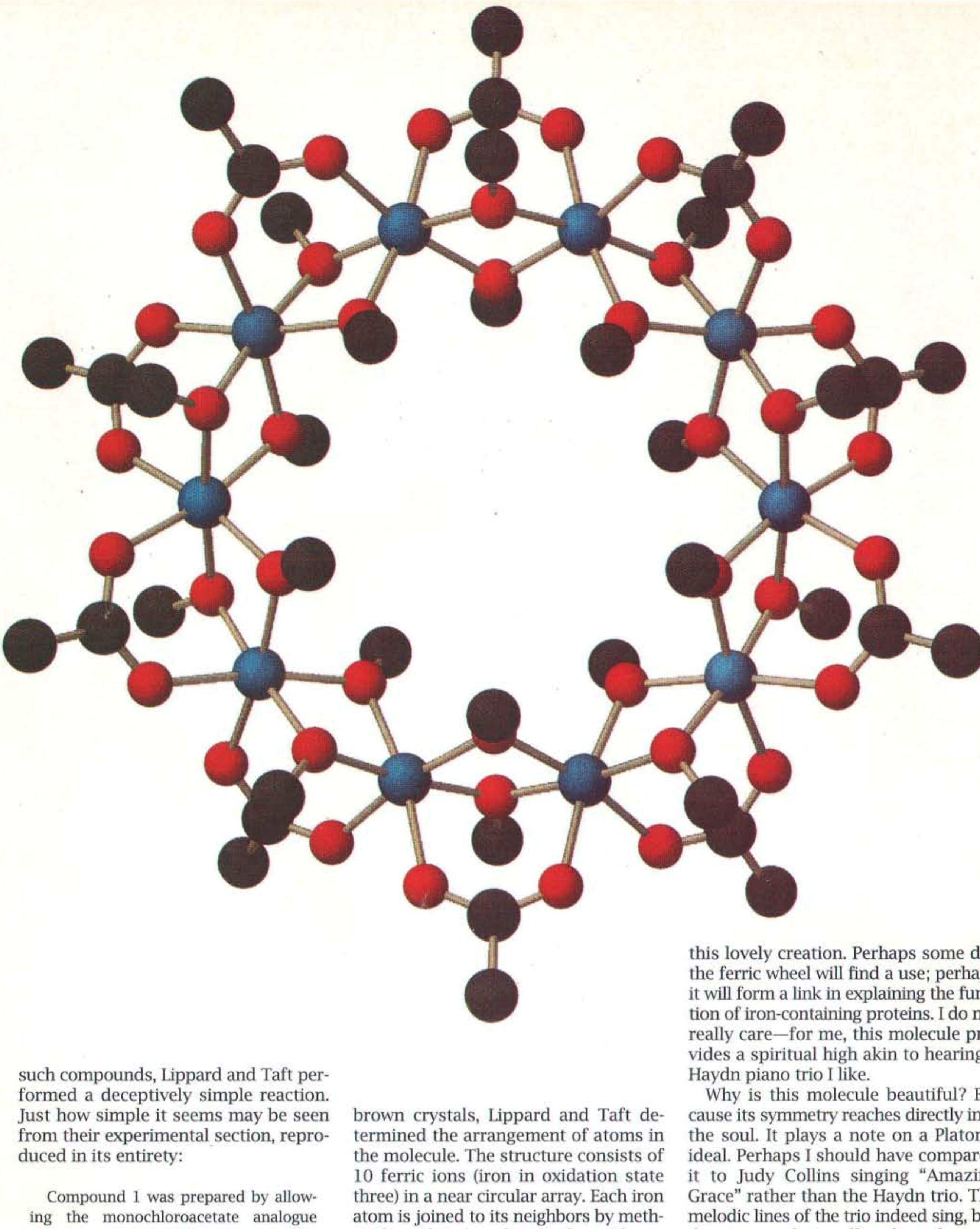
The synthesis of thienamycin is a building process, proceeding from simple pieces to the complex goal. It shares many features with architecture. For instance, a necessary intermediate structure may be more complicated than either the beginning or end; think of scaffolding. Chemical synthesis is a local defeat of entropy, just as our buildings and cities are. The analogy to architecture is so strong that one forgets how different, how marvelous, this kind of construction is. In a flask there may be 10^{23} molecules, moving rapidly, colliding often. Hands off, following only the strong dictates of thermodynamics, they proceed to shuffle their electrons, break and make bonds, do our bidding. If we're lucky, 99 percent of them do.

Chemists can easily calculate, given a certain number of grams of starting material, how much product one should get. That is the theoretical yield. The actual amount obtained is the experimental yield. There is no way to get something out of nothing but many ways to get less than you theoretically could. One way to achieve a 50-percent yield is to spill half the solution on the floor. This will impress no one. But even if you perform each transfer as neatly as possible, nature may not give you what you desire but instead transform 70 percent into black gunk. This is also not impressive, for it does not demonstrate control of mind over matter. Experimental yields are criteria not only of efficiency, essential to the industrial enterprise, but also of elegance and control.

There is more, much more, to say about the planned organic synthesis. But let me go on to my second case study: the ferric wheel.

Stephen J. Lippard and Kingsley L. Taft of the Massachusetts Institute of Technology synthesized the ferric wheel, also known as $[Fe(OCH_3)_2(O_2CCH_2Cl)]_{10}$. They discovered this exquisite molecule while studying model molecules for inorganic reactions that occur in biological systems. For instance, a cluster of iron and oxygen atoms is at the core of several important proteins, such as hemerythrin, ribonucleotide reductase, methane monooxygenase and ferritin (not household words these, but essential to life).

In the course of their broad attack on



such compounds, Lippard and Taft performed a deceptively simple reaction. Just how simple it seems may be seen from their experimental section, reproduced in its entirety:

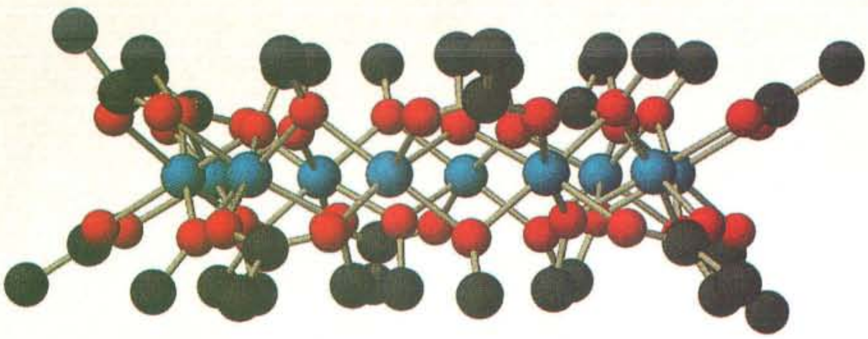
> Compound 1 was prepared by allowing the monochloroacetate analogue of basic iron acetate, $[Fe_3O(O_2CCH_2Cl)_6(H_2O)_3](NO_3)$ (0.315 g, 0.366 mmol), to react with 3 equiv of $Fe(NO_3)_3{\cdot}9H_2O$ (0.444 g, 1.10 mmol) in 65 mL of methanol. Diffusion of ether into the green-brown solution gave a yellow solution, from which both gold-brown crystals of 1 and a yellow precipitate deposited after several days.

Using x-ray diffraction on the gold-brown crystals, Lippard and Taft determined the arrangement of atoms in the molecule. The structure consists of 10 ferric ions (iron in oxidation state three) in a near circular array. Each iron atom is joined to its neighbors by methoxide and carboxylate bridges, "forming a molecular ferric wheel," to quote its makers.

No one will deny the visual beauty of this molecule. It does not have the annual sales of Primaxin, estimated to be \$500 million. On the contrary, it probably cost the U.S. taxpayer several thousand dollars to make it. But I do not know a single curmudgeonly chemist who would not respond positively to this lovely creation. Perhaps some day the ferric wheel will find a use; perhaps it will form a link in explaining the function of iron-containing proteins. I do not really care—for me, this molecule provides a spiritual high akin to hearing a Haydn piano trio I like.

Why is this molecule beautiful? Because its symmetry reaches directly into the soul. It plays a note on a Platonic ideal. Perhaps I should have compared it to Judy Collins singing "Amazing Grace" rather than the Haydn trio. The melodic lines of the trio indeed sing, but the piece works its effect through counterpoint, the tools of complexity. The ferric wheel is pure melody.

Were we to write out the synthesis of the ferric wheel, there would be but a single arrow, from the iron chloroacetate and ferric nitrate to the product. This is a very different type of synthesis—the product essentially self-assembles to its final glory. When I see such a process, much more typical of inorgan-



FERRIC WHEEL, formally known as $[Fe(OCH_3)_2(O_2CCH_2Cl)]_{10}$, exemplifies the ideal in chemistry. The molecule consists of iron (*blue*), oxygen (*red*) and carbon (*black*). To highlight the symmetry, the chlorine and hydrogen atoms are not shown.

ic systems than organic ones, I immediately wonder what I'm missing. The Swedish chemist Sture Forsén has aptly expressed the frustration in not being able to observe the intermediate stages of a reaction:

> The problem facing the scientist has been compared with that of a spectator of a drastically shortened version of a classical drama—"Hamlet," say—where he or she is only shown the opening scenes of the first act and the last scene of the finale. The main characters are introduced, then the curtain falls for change of scenery, and as it rises again we see on the scene floor a considerable number of "dead" bodies and a few survivors. Not an easy task for the inexperienced to unravel what actually took place in between.

Wheels, ferric or ferris, don't really self-assemble in one fell swoop. It remains for us to learn in the future how those bridges and irons come together.

Some chemists, especially those who practice the mentally demanding, intellectually exhilarating many-step, planned synthesis of the thienamycin type look askance at one-step self-assembly. Such one-fell-swoop syntheses are especially common in solid-state chemistry, in the formation of materials extended infinitely in one, two or three dimensions. The high-temperature superconductors are a good example of molecules made just this way. Their synthesis does not appear to show control of mind over matter. It looks like magic.

I exaggerate, but this is one strand of thought in the community. If I could corner my straw-man scoffer at self-assembly, typically an organic chemist, and engage him or her in a Socratic dialogue, I would begin with the question "When have you made any diamond for me lately?" Diamond is a beautifully simple three-dimensional structure (natural!). It contains in it six-membered rings, the bread-and-butter of organic chemistry. Such rings of carbon atoms are easy to make in a discrete molecule. But diamond can be made only by techniques organic chemists find unsporting, by discharges forming a plasma in methane or by pressing graphite.

Organic chemists are masterful at exercising control in zero dimensions. To one piece of carbon, perhaps asymmetric, they add another piece. Slowly, painstakingly, a complex edifice emerges. (Thienamycin is pretty simple compared to what you can do today.) One subculture of organic chemists has learned to exercise control in one dimension. These are polymer chemists, the chain builders. Although they may not have as much honor in organic chemistry as they should, they do earn a good bit of money.

But in two or three dimensions, it's a synthetic wasteland. The methodology for exercising control so that one can make unstable but persistent extended structures on demand is nearly absent. Or to put it in a positive way—this is a certain growth point of the chemistry of the future.

Syntheses, like human beings, do not lend themselves to typology. Each one is different; each has virtues and shortcomings. From each we learn. I will stop, however reluctantly, with primaxin and the ferric wheel and turn to some general questions they pose, especially about the natural and the unnatural.

Two paradoxes are hidden in the art of synthesis. The first is that the act of synthesis is explicitly human and therefore unnatural, even if one is trying to make a product of nature. The second is that in the synthesis of ideal molecules, where doing what comes unnaturally might seem just the thing, one sometimes has to give in to nature. Let me explain in the context of the two syntheses I have just discussed.

Imipenem, one component of the successful Merck antibiotic, is made from thienamycin. The thienamycin is natural, to be sure, but an economic and chemical decision dictated that in its commercial production thienamycin be made synthetically.

There is no doubt in this case that the natural molecule served as an inspiration for the synthetic chemists. But, of course, they did not make thienamycin in the laboratory the way it is made by the mold. The organism has its own intricate chemical factories, enzymes shaped by evolution. Only recently have we learned to use genetic engineering to harness those factories, even whole organisms, for our own purposes.

We have grown proficient at simpler, laboratory chemistries than those evolved by biological organisms. There is no way that Christensen and his team would set out to mimic a mold enzyme in detail. They did have confidence that they could carry out a very limited piece of what the lowly mold does, to make thienamycin, by doing it differently in the laboratory. Their goal was natural, but their process was not.

To make thienamycin, Christensen and his co-workers used a multitude of natural and synthetic reagents. For instance, one of their transformations—the synthesis of compound 3 [*see box on page 42*]—uses a magnesium compound, $(CH_3)_3CMgCl$, known as a Grignard reagent. Magnesium compounds are abundant in nature (witness Epsom salt and chlorophyll). But the reagent in question, a ubiquitous tool of the synthetic chemist, was concocted by Victor Grignard some time around the turn of the century. The creation of compound 3 also requires treatment with hydrochloric acid and ammonium chloride, both natural products. (Your stomach has a marginally lower concentration of hydrochloric acid than that used in this reaction, and ammonium chloride is the alchemist's sal ammoniac.) But even though these molecules occur in nature, they are far easier to make in a chemical plant.

Because everything in the end does come from the earth, air or water, every unnatural reagent used in the synthesis ultimately derives from natural organic or inorganic precursors. The very starting material in the synthesis of imipenem is an amino acid, aspartic acid.

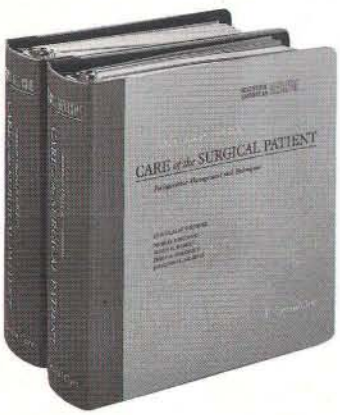
Now consider the most unnatural and beautiful ferric wheel. It was made simply by reacting two synthetic molecules, the iron monochloroacetate and ferric nitrate, in methanol, a natural solvent. The methanol was probably made synthetically; the two iron-containing reagents derive from reactions of iron metal, which in turn is extracted from iron ores. And the final wrinkle is the method of assembly: the pieces of the molecule seem to just fall into place (self-assembly). What could be more nat-

SYSTEMS THAT CHANGE RIGHT ALONG WITH IT."

For unrivaled scope of coverage, technical excellence, and attention to detail, turn to the new and improved...

CARE *of the* SURGICAL PATIENT

Designed and published in cooperation with the American College of Surgeons.



"...A much needed comprehensive overview of surgical care... highly recommended." —Annals of Surgery

"A veritable who's who of nationally and internationally recognized surgical experts. Remarkably consistent and eminently readable throughout." —New England Journal of Medicine

"Truly state of the art.... Should remain the definitive work for many years to come." —JAMA

An unequaled professional reference. Nearly 100 distinguished contributors have produced 1,500 pages of the most authoritative information available today on perioperative management and techniques. And it's updated quarterly to keep in step with your need for up-to-date information.

CARE *of the* SURGICAL PATIENT provides two vital benefits:

Saves you Money! Comparable in price to many single journal subscriptions with a scope of coverage equal to a dozen journals and textbooks.

Saves you Time! You can spend 30 minutes or more leafing through stacks of books and journals or you can get up-to-date answers in 30 seconds or less with CARE *of the* SURGICAL PATIENT.

EDITORIAL BOARD

Douglas W. Wilmore, M.D.
Harvard Medical School, Brigham and Women's Hospital

Murray F. Brennan, M.D.
Memorial Sloan-Kettering Cancer Center

Alden H. Harken, M.D.
University of Colorado Health Sciences Center

James W. Holcroft, M.D.
University of California, Davis Medical Center

Jonathan L. Meakins, M.D., D.Sc.
McGill University Faculty of Medicine, Royal Victoria Hospital

Three Important New Benefits Added for 1993

NEW

Four updates per year to provide more clinical information on important developments on an even more timely basis.

NEW

Individual subscribers may earn up to 30 hours of Category 2 CME credit per year through the American College of Surgeons at no additional cost.

NEW

Technique supplements, periodically included in the update packages, bring you up-to-date on new evolving techniques.

PLUS

- Hundreds of beautiful illustrations and photographs (many in full color), and helpful tables and charts.
- Practical details on clinical problems.
- Answers to your questions about diagnostic procedures, therapeutic maneuvers, perioperative medications.
- Complete coverage of the ICU, emergency room, trauma, infections, elective complications.
- Concise diagnosis and management algorithms.
- A News Bulletin summarizing important advances and a complete, easily referenced 50-page index with each update.

CARE *of the* SURGICAL PATIENT is available in two looseleaf volumes containing over 1,500 pages *plus* the first year's updates FREE. (The first update is inserted in the binders prior to shipment.) If you are not completely satisfied, simply return the volumes to us within 30 days for a complete refund.

For price and availability contact: W.H. Freeman and Company Limited, 20 Beaumont Street, Oxford OX1 2NQ, England Tel: 0865 726975.

ural than letting things happen spontaneously, giving in to the strong dictates of entropy?

It is clear that in the unnatural making of a natural molecule (thienamycin) or of an unnatural one (the ferric wheel), natural and synthetic reagents and solvents are used in a complex, intertwined theater of letting things be and of helping them along. About the only constant is change, transformation.

We may still wonder about the psychology of chemical creation. Which molecules should we expend our energies in making? Isn't there something inherently better in trying to make the absolutely new?

Four beautiful polyhedra of carbon have piqued the interest of synthetic organic chemists during the past 40 years: tetrahedrane (C_4H_4), cubane (C_8H_8), dodecahedrane ($C_{20}H_{20}$) and buckminsterfullerene (C_{60}). Cubane is quite unstable because of the strain imposed at each carbon. (In cubane the angle between any three carbon atoms is 90 degrees, but each carbon would "prefer" to form angles of 109.5 degrees with its neighbors.) C_{60} is also somewhat strained because of both its nonplanarity and its five-membered rings. Tetrahedrane is particularly unstable. One has to create special conditions of temperature and solvent to see it; even then, the parent molecule has not yet been made, only a "substituted derivative," in which hydrogen is replaced by a bulky organic group.

As far as we know, tetrahedrane, cubane and dodecahedrane do not exist naturally on the earth. C_{60} has been found in old soot and a carbon-rich ancient rock, shungite. It may turn up elsewhere. Be that as it may, all four molecules were recognized as synthetic targets at least 20 years, in some cases 50 years, before they were made. Some of the best chemists in the world tried to make them and failed. The syntheses of cubane and especially dodecahedrane were monumental achievements in unnatural product chemistry.

C_{60} was different. The pleasing polyhedral shape was first noted by some theoreticians. Their calculations indicated some stability; such indications as the theoreticians had at their command were sometimes unreliable. These theoreticians' dreams were ignored by the experimentalists and by other theoreticians. It is sometimes difficult to see the shoulders of the giants we stand on when we are looking so intently ahead. I myself have suggested a still unsynthesized metallic modification of carbon, different from diamond or graphite, and even though I have substantially more visibility among chemists than the proposers of buckminsterfullerene, no one has paid much attention to my pipe dream either, probably for good reason. We see what we want to see.

One organic chemist I know, a very good one, Orville L. Chapman of the University of California at Los Angeles, independently thought up the structure and devoted much time to the planned, systematic making of C_{60}. After all, this was a "simple" molecule, not an extended material like the repeating lattice of carbon atoms that make up a diamond. So it should be possible to make it. Despite persistent efforts over a 10-year period, Chapman and his students failed in their effort.

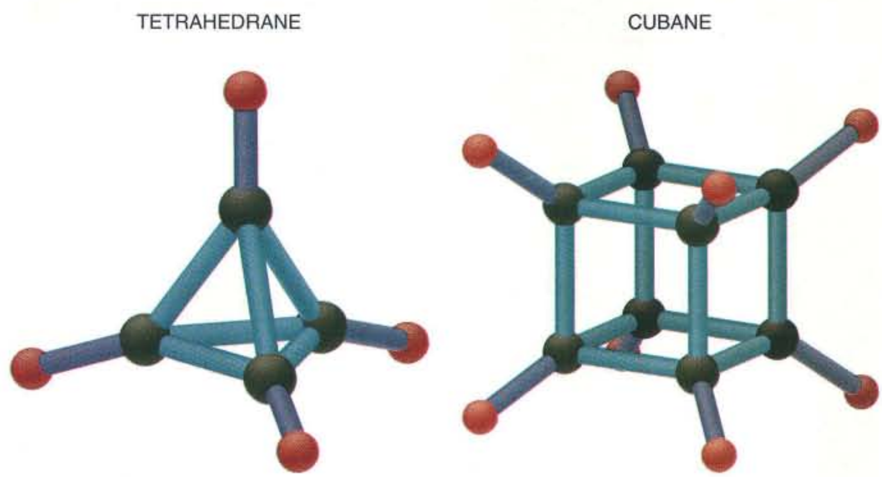
The first evidence, indirect but definitive, for C_{60} was obtained from a very different branch of our science, physical chemistry. The credit for the discovery belongs properly to Richard E. Smalley and Robert F. Curl of Rice University and Harold W. Kroto of the University of Sussex. They obtained hard evidence for tiny amounts of C_{60} in the gas phase, assigned the molecule its name and, more important, deduced its structure. Did they make it? Absolutely. It did not matter to me or to other believers in their evidence that they had made "just" 10^{10} molecules instead of the 10^{20} we need to see in a tiny crystal. But there were doubters, many I suspect, in the organic community. One wanted to see the stuff.

Grams of buckminsterfullerene were provided by a synthesis by Donald R. Huffman of the University of Arizona and Wolfgang Krätschmer and Konstantinos Fostiropoulos of the Max Planck Institute for Nuclear Physics in Heidelberg. Striking a carbon arc in a helium atmosphere (which is what they did) is about as unsporting as firing a laser at graphite (the Smalley-Kroto-Curl synthesis). But it certainly makes plenty of C_{60}, enough of the molecule to determine its structure by typical organic methods, enough to convince any chemist that it has the soccer-ball structure ["Fullerenes," by Robert F. Curl and Richard E. Smalley; SCIENTIFIC AMERICAN, October 1991].

I think many chemists wished C_{60} had been made in a planned, unnatural way. I am happy that—just to make the world slightly less rational than we would like it to be—it was made in a serendipitous way.

Serendipity—a word invented by Horace Walpole—has come to mean "a discovery by chance." Yet whether it is a chemical synthesis or a Japanese master potter piling organic matter around the ceramic objects in his Bizen kiln, chance favors the prepared mind. You need to have the knowledge (some call it intuition) to vary the conditions of striking the arc or the arrangement of the leaves in the kiln just so. You need to have the instruments and intuition to deduce structure from a few fuzzy lines in a spectrum and to reject false leads. And you need to have the courage to shatter a vase that didn't come out right and to learn from one firing what to do in the next.

Many chemical syntheses, even if part of a grand design, proceed by steps that are serendipitous. One wants to link up a bond here, but it doesn't work. So one follows a hunch, anything but the codified scientific method. One knows that if a reaction works, one can construct a

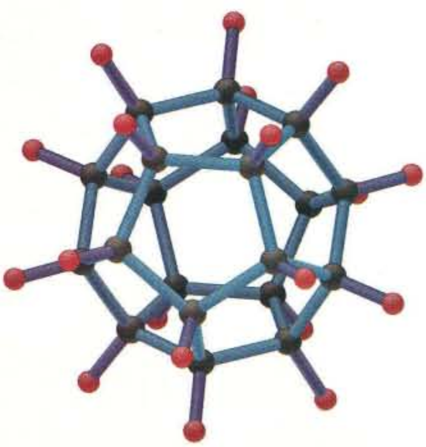


FOUR POLYHEDRA based on carbon were recognized decades ago as targets for chemists to synthesize. Buckminsterfullerenes were discovered in 1985 and were then found to occur naturally on the earth. Tetrahedrane, the simplest of the struc-

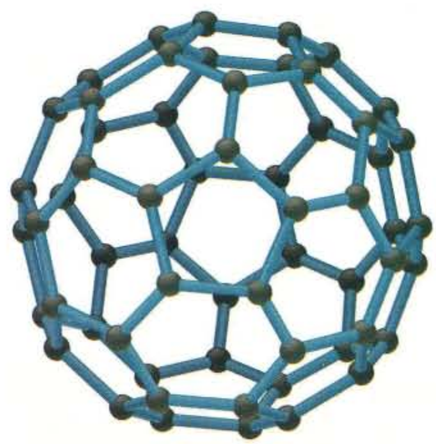
rationalization for it—an argument spiffy clean enough to make an impression on one's colleagues. Eventually one can make the damned reaction work if it is a necessary step in the design.

Because chance also operates to foil every design, it is almost certain that in the course of any planned synthesis there will be a step that will not work by any known process. So a new one will be invented, adding to the store of the chemists, aiding others around the world facing the same problem. Some synthetic chemists—for instance, E. J. Corey of Harvard University, a grand master of the art—have a special talent for not only making interesting molecules but also using the opportunity of the synthesis to introduce a brilliant, unprecedented methodology, applicable to other syntheses.

DODECAHEDRANE



BUCKMINSTERFULLERENE



tures, has not yet been synthesized. Tetrahedrane, cubane and dodecahedrane consist of carbon atoms (*black*) and hydrogen atoms (*red*). Buckminsterfullerenes are made solely from carbon.

When the synthesis is planned, be its aim a natural or unnatural molecule, we suppress the aleatory nature of the enterprise. We want to project an image of mind over matter, of total control. When the molecule made is unanticipated, as the ferric wheel was, we find it very difficult to hide the workings of chance. But hazard—to use the meaning that is dominant in the French root of our word, and secondary in ours—plays an unrecognized and enlivening role in all synthesis.

Let us return to nature and our struggles to emulate it. Or surpass it. Can we make substances that have properties superior to those found in nature? I say "yes" while recognizing that the phrase "superior to nature" is patently value laden and anthropocentric and should immediately evoke ecological concerns.

There is nylon instead of cotton in fishing nets, nylon instead of silk in women's stockings. No one, least of all Third World fishermen, will go back to the old nets. Some people may go back to silk stockings, but they will only be the rich, out to impress. There are new chemical materials and new combinations of old materials for dental restorations. They make a world of difference to older people in this world, and their benefit cannot be dismissed.

Yet the thought that we can do better than nature is provocatively arrogant. As we have attempted to improve on nature (while failing to control the most natural thing about us, our drive to procreate), we have introduced so many transformations and in such measure that we have fouled our nest and intruded into the great cycles of this planet. We must face the reality that natural evolution proceeds far too slowly to cope with our changes. This is a concern that, just as much as utility, should guide the industrial-scale syntheses of the future.

I want to touch on another kind of human arrogance implicit in the intellectual drama of synthesis. A French chemist, Alain Sevin, has put it well:

> The incredible richness and fantasy of Nature is an act of defiance to Man, as if he had to do better in any domain. Flying faster than birds, diving deeper than whales.... We are Promethean characters in an endless play which now is in its molecular act.

We are driven to transform. We have learned to do it very well. But this play is not a comedy.

Were chemical synthesis in search of a single icon, the outstretched hand of Prometheus bringing fire to humanity would serve well. Prometheus, a name meaning "forethought," represents the element of design, the process of fruitfully taking advantage of chance creation. Fire is appropriate because it drives transformation. The hand of Prometheus is the symbol of creation—the hand of God reaching to Adam in Michelangelo's fresco, the hands in contentious debate in Dürer's *Christ among the Doctors,* the infinite variety of hands that Rodin sculpted. Hands bless, caress and hide, but most of all, they shape.

The sculptor's art itself mimics the complexity of motion of a chemist across the interface between natural and unnatural. Rodin, in his human act of creation, sketches, then shapes by hand (with tools) an out-of-scale yet "realistic" artifact, a sculpture of a hand, out of materials that are synthetic (bronze) but that have natural origins (copper and tin ores). He uses a building process (maquettes, a cast) that is complex in its intermediate stages. The sculptor creates something very real, whose virtue may reside in calling to our minds the ideal.

Margaret Drabble has written that Prometheus is "firmly rooted in the real world of effort, danger and pain." Without chemical synthesis, there would be no aspirin, no cortisone, no birth-control pills, no anesthetics, no dynamite. The achievements of chemical synthesis are firmly bound to our attempt to break the shackles of disease and poverty. In search of an ideal, making real things, the mind and hands engage.

FURTHER READING

SYNTHESIS. R. B. Woodward in *Perspectives in Organic Chemistry.* Edited by A. R. Todd. Interscience, 1956.

A STEREOCONTROLLED, ENANTIOMERICALLY SPECIFIC TOTAL SYNTHESIS OF THIENAMYCIN. T. N. Salzmann, R. W. Ratcliffe, F. A. Bouffard and B. G. Christensen in *Philosophical Transactions of the Royal Society of London,* Series B, Vol. 289, No. 1036, pages 191–195; May 16, 1980.

MOLECULES. Peter W. Atkins. Scientific American Library, 1987.

THE LOGIC OF CHEMICAL SYNTHESIS. E. J. Corey and Xue-Min Cheng. John Wiley and Sons, 1989.

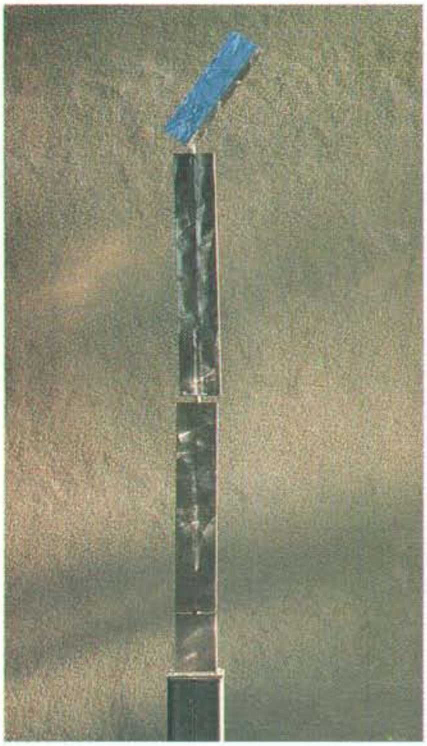
SYNTHESIS AND STRUCTURE OF $[Fe(OMe)_2(O_2CCH_2Cl)]_{10}$, A MOLECULAR FERRIC WHEEL. Kingsley L. Taft and Stephen J. Lippard in *Journal of the American Chemical Society,* Vol. 112, pages 9629–9630; December 19, 1990.

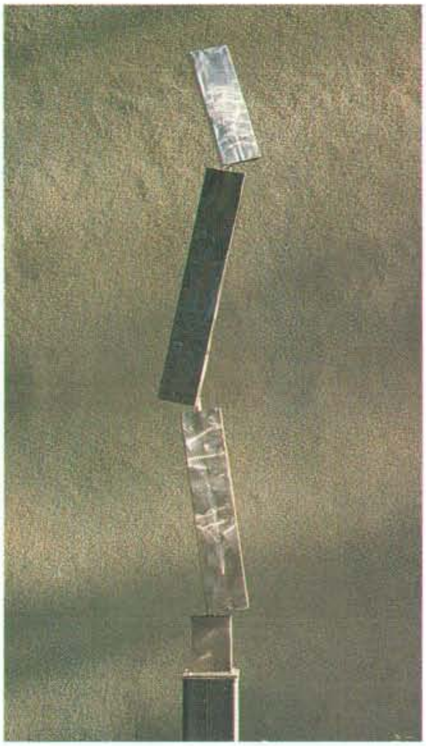
THE ORGANIC CHEMISTRY OF DRUG DESIGN AND DRUG ACTION. Richard B. Silverman. Academic Press, 1992.

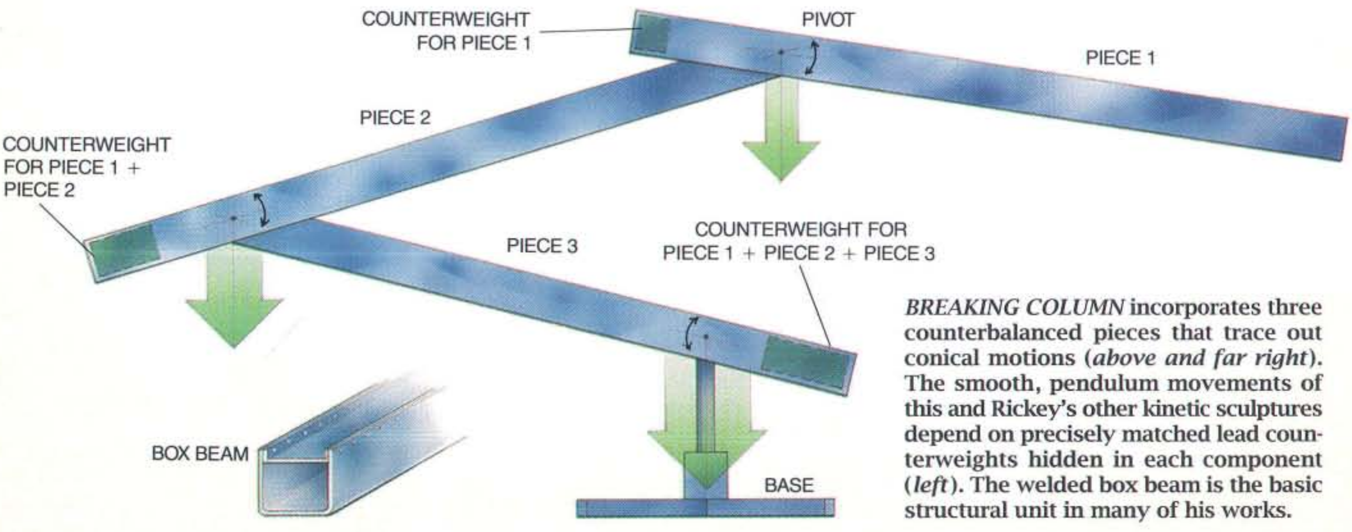
A Technology of Kinetic Art

Delicate interplay of weights and balances choreographs the author's sculptures so that the gentlest gusts of air set their parts in motion

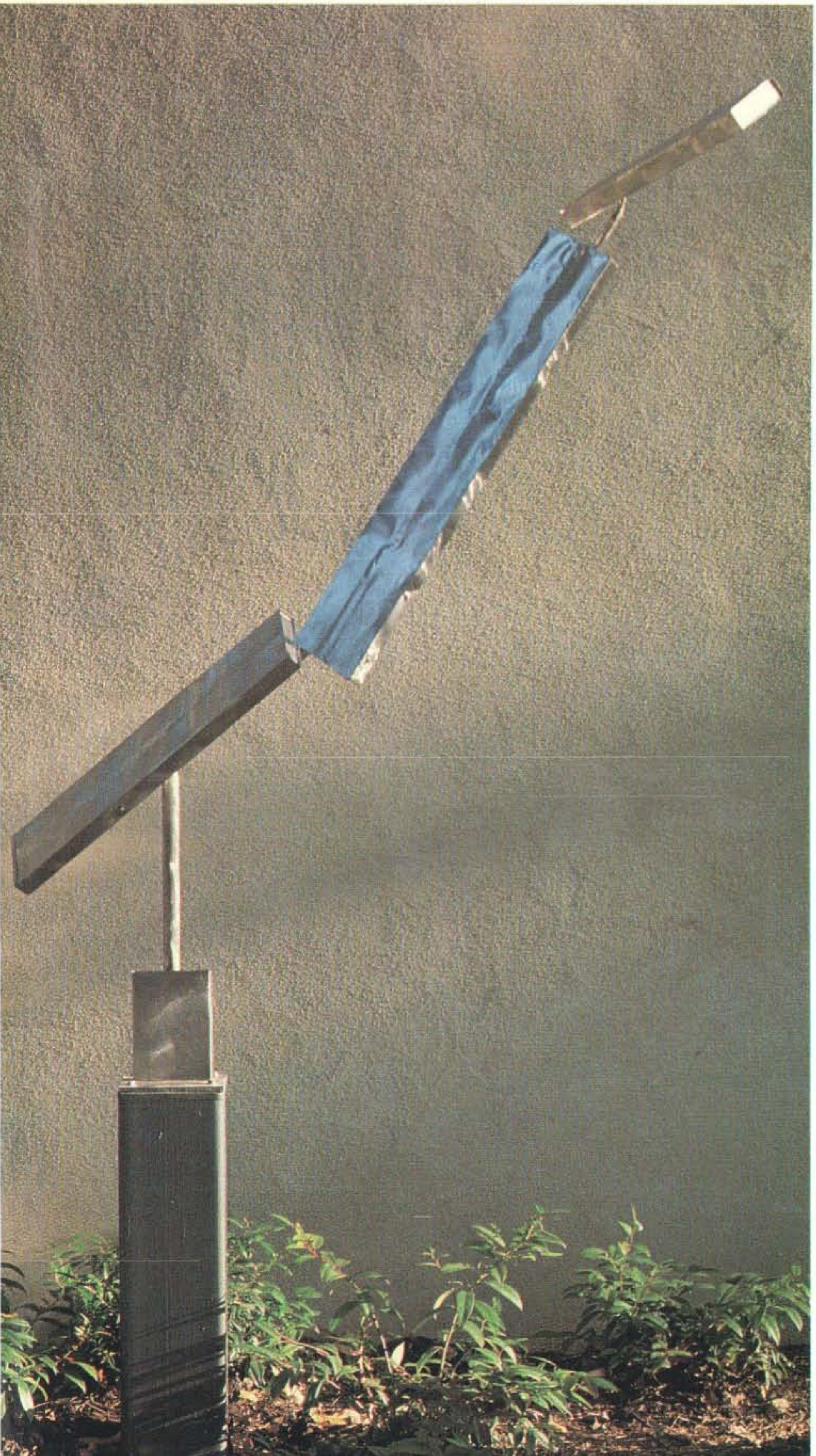
by George Rickey







BREAKING COLUMN **incorporates three counterbalanced pieces that trace out conical motions (*above and far right*). The smooth, pendulum movements of this and Rickey's other kinetic sculptures depend on precisely matched lead counterweights hidden in each component (*left*). The welded box beam is the basic structural unit in many of his works.**



GEORGE RICKEY has spent the past 40 years developing and refining his kinetic sculptures. He was born in South Bend, Ind., in 1907; in 1913 he moved to Scotland. While at Balliol College at the University of Oxford, Rickey surreptitiously attended drawing classes at the Ruskin School. After army service, he attended the Institute of Design in Chicago and began exploring the possibilities of mobile sculpture. Since 1960 Rickey has worked and lived in upstate New York. Retrospective exhibitions of his art have appeared at the National Gallery, Berlin (1973), the Guggenheim Museum (1979), Glasgow (1982) and Gropius Bau, Berlin (1992).

My artistic ideas, like the motions of my sculptures themselves, have unfolded in a casual and unexpected manner. By the time I reached 22 years of age, I had run errands for a fashion scout, taught English in a language school and exhausted my childhood savings to attend the Lhote Academy, a cubist-influenced painting school in Paris. Not until 13 years later, when I served in the U.S. Army Air Corps during World War II, did I discover a mechanical impulse in my fingers and in my head. In the course of working with industrial tools and military hardware, I started toying with the notion of dynamic constructions that were neither static sculptures nor mobiles in the traditional sense. After the war, I set about transforming those ideas into tangible reality. I exhibited the resulting works for the first time in 1953, when I was 46.

Since then, I have continued to explore the aesthetic possibilities of kinetic sculpture. In my recent pieces, I hang one moving element—often a simple geometric form like a rod or a cube—onto another, to produce double-jointed or triple-jointed devices. An almost exactly even distribution of weight allows them to respond to the slightest wafts of air, tracing slow, unpredictable but expressive convolutions.

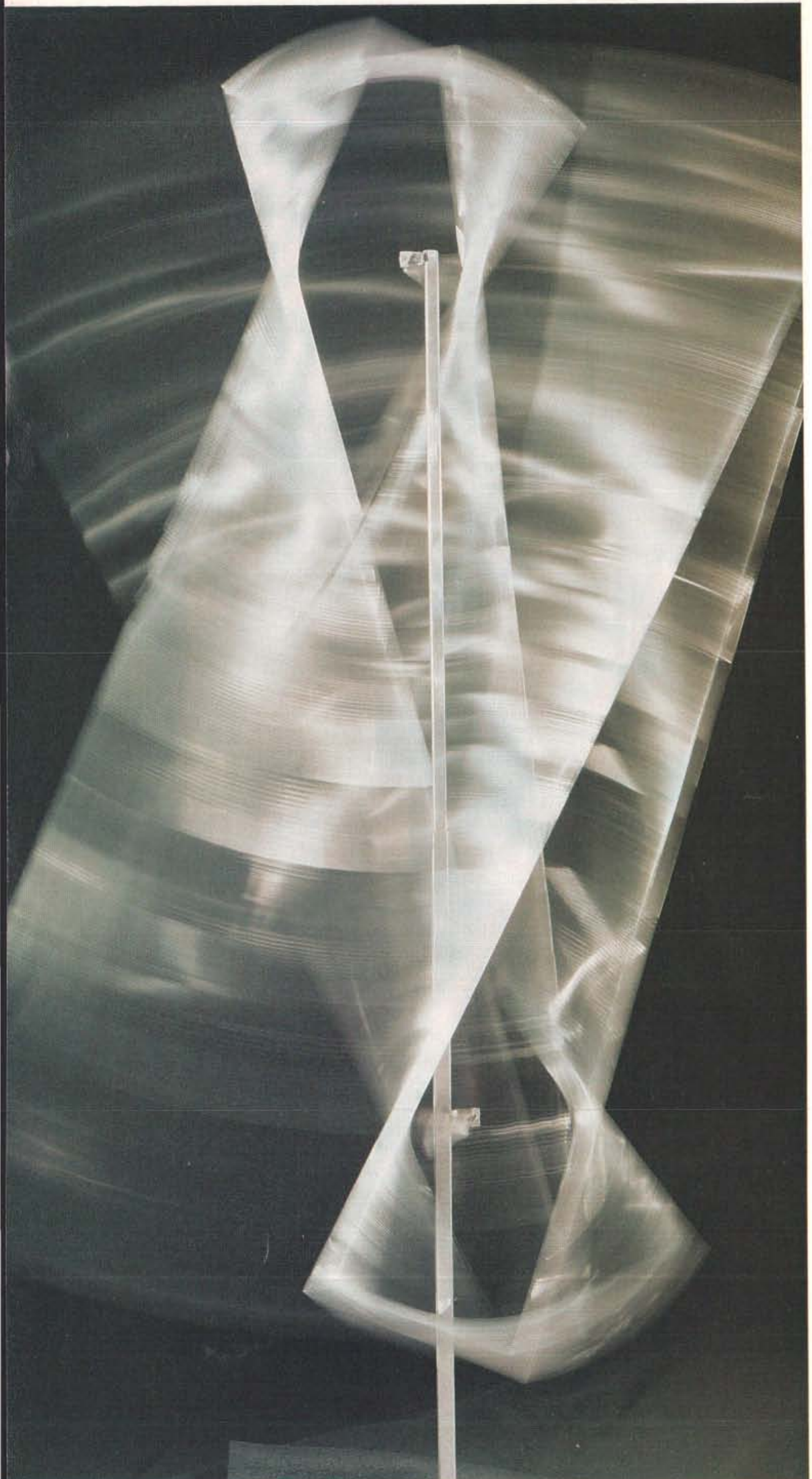
At the most elementary level, my works function as modified pendulums. I must credit my grandfather, a clockmaker in Athol, Mass., for directing my thinking along these lines. When I was five, he let me disassemble a Swiss clock that used wooden gears. I did it, but I am still chagrined that I could not put it together again.

I learned about so-called compound pendulums some 40 years after flunking my grandfather's test. Unlike the simple pendulum in a clock, a compound pen-

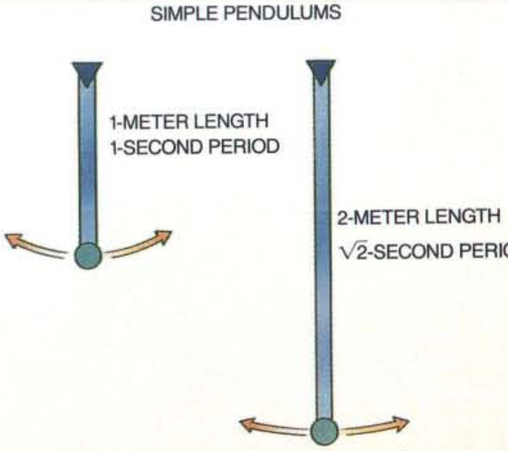
dulum carries weights both above and below the pivot, or fulcrum, on which the main rod swings. For a simple pendulum, the time required to complete one swing depends only on the length of the rod. For a compound pendulum, however, one can control the period of the swing by changing the distribution of weights; the more weight above the fulcrum, the slower the swing. When the product of weight and distance above the fulcrum equals the product of weight and distance below, the period stretches to infinity, and the compound pendulum becomes unstable. In another shape, such a pendulum takes on a familiar guise: the wheel.

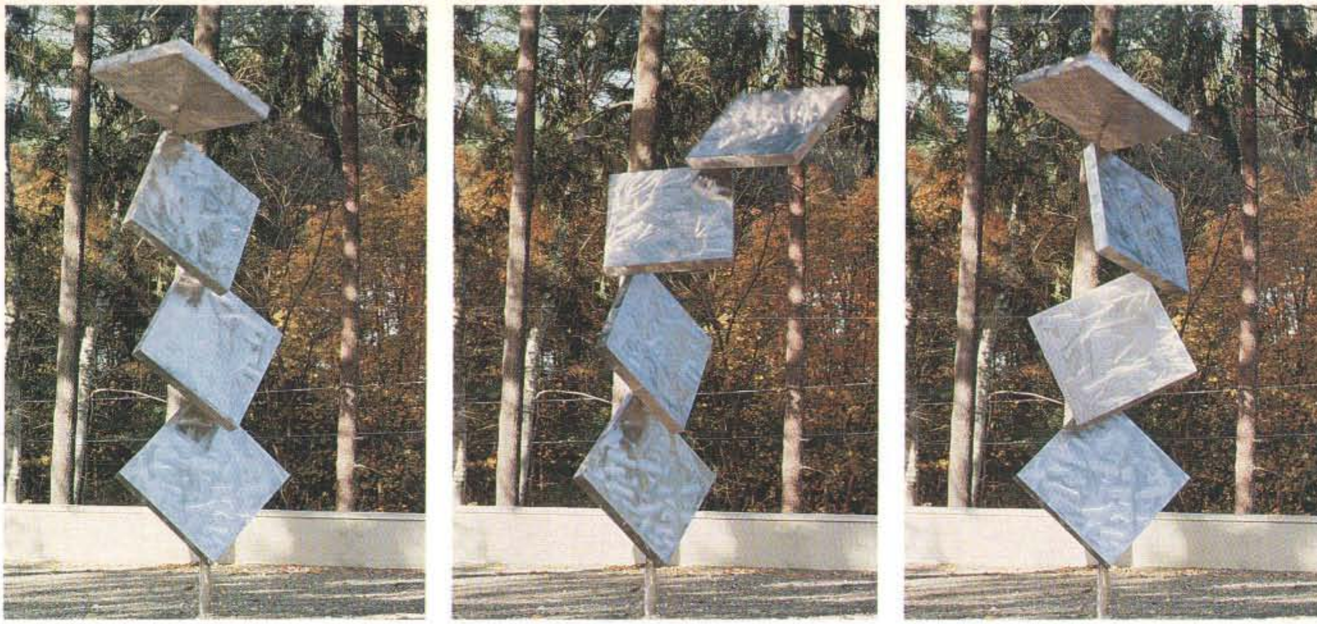
I play with arrangements of weight and distance that lie on the brink of instability to achieve slow, lyrical motions; a one-way swing in one of my sculptures may take 20 seconds or longer. Such arrangements possess another noteworthy attribute: when the distribution of weight approaches instability, the forces needed to set the pendulum in motion diminish to almost nothing. A breeze that softly stirs leaves can also move a sculpture weighing 500 kilograms. Control of weight and balance—and also of time—gives me a means of expression comparable to color for a painter or sound for a composer.

A great deal of my effort has been



COMPOUND PENDULUMS form the basis of Rickey's sculptures. The time required for one swing of a simple pendulum (*below*) depends only on its length; the period of a compound pendulum (*right*) is determined by the product of weight times distance, above and below the pivot. Properly weighted pendulums can swing slowly in response to minuscule forces. A time-lapse view of *Open Triangles One Up One Down* (*left*) reveals its subtle swaying motions.



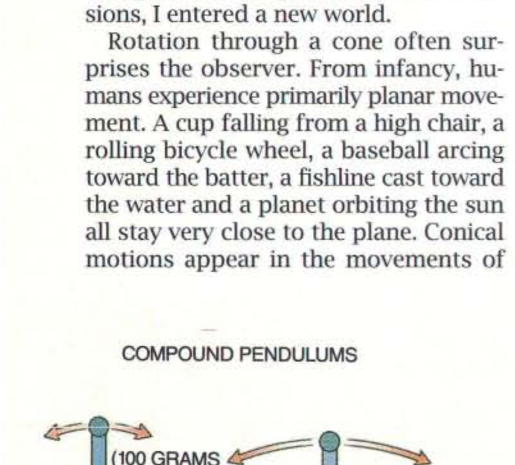


LINKED SQUARES are the result of Rickey's recent sculptural experiments. Each element in *Column of Four Squares Eccentric Gyratory III* is weighted and mounted on an angled bearing. The bottom shaft carries all the weight to the stable base.

directed toward escaping from the slicing confinement of motion within a plane. Some 20 years ago, while trying to expand the possibilities of expressive movement, I experimented with changing the axis of rotation at the pendulum's fulcrum to 45 degrees rather than the usual 90. The resulting path traced out a cone instead of the flat arc familiar to anyone who has ever been on a swing. In breaking out of two dimensions, I entered a new world.

Rotation through a cone often surprises the observer. From infancy, humans experience primarily planar movement. A cup falling from a high chair, a rolling bicycle wheel, a baseball arcing toward the batter, a fishline cast toward the water and a planet orbiting the sun all stay very close to the plane. Conical motions appear in the movements of a ship, which rocks and rolls, pitches and yaws, rises and falls, and moves backward and forward. These movements are well known to sailors, fliers and the seasick. I discovered they could also be quite beautiful.

COMPOUND PENDULUMS



My investigations of complex, jointed constructions have brought me face to face with a variety of technical problems. I have learned that sensitive and controlled responses require that a sculpture have a carefully adjusted distribution of mass. The accumulated weight of a sequence of linked parts (the rigid beams, bearings and counterweights) must exactly balance the counterweight of the innermost unit. The sculpture cannot be completed unless that innermost unit is large enough to accommodate a sufficiently heavy counterweight. Conversely, the space available for inserting weights inside the heavy inner unit determines the maximum dimensions of the outermost one.

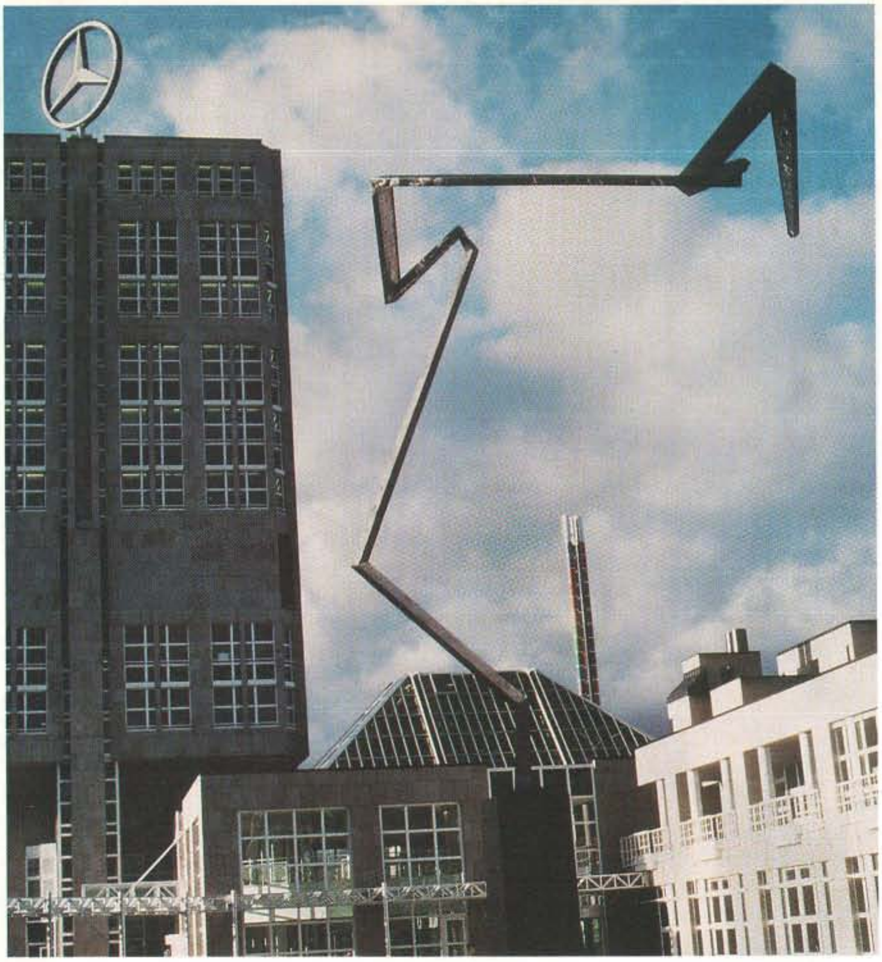
Rigidity of the structural beams is as critical as precise counterbalancing; both are essential to the genesis of works whose leisurely movements force the viewer to wait and wonder. At the same time, I want the lightest units in my kinetic sculptures to appear indistinguishable from the heaviest. I have come to rely on a simple box beam that I construct by spot welding together two U-shaped steel pieces.

My designs demand also that I keep friction in the mobile joints to a minimum; too much drag would cause the sculpture's suspended wanderings to die out prematurely. Knife-edge bearings, such as are used in laboratory scales, have the advantage of extremely low friction, but they can wear down rapidly or be damaged by rain and heavy winds. In works designed for outdoors, I use ball bearings, which tolerate exposure to inclement weather. Ball bearings need grease, however, whose viscosity can dampen motions in very light breezes.

Because of the nature of the balances, I am forced to build a sculpture from the outside in. For a two-part work, I balance the outer, lighter unit and attach it to its inward neighbor, which must carry enough counterweight to balance both components. I then insert the shaft, located at the balance point of this assembly, into a bearing that can carry the sculpture's total weight to ground, wall or ceiling.

In 1985 I was invited to contribute a public work in Berlin to celebrate that city's 750th birthday in 1987. My sculpture would be placed in Breitscheidplatz, near the Memorial Church in downtown West Berlin. This opportunity prompted me to think about extending my techniques even further. I had already hinged one linear part onto its neighbor to make an elbow. Could I now hinge a cone on a cone, like a wrist or a shoulder joint? Vertebrate animals do it; perhaps, I thought, so could I.

Pushing onward into this new territory, I mounted one angular component onto another. Each part rotated on a shaft protruding at 45 degrees to the long dimension; the entire assembly was mounted on a zigzag support. The parts would swing in the wind through



PUBLIC PROJECTS have provided an impetus for further artistic explorations. A sculpture built for the 750th Jubilee of the founding of Berlin (*above*), now located near Stuttgart, has two zigzag-shaped pieces free to move along conical paths. A subsequent work designed for a theater in Rotterdam (*below*) posed a greater technical challenge because each member of the triad has three moving parts.

virtual cones, barely missing bumping into the support and into one another. I adjusted the weights to yield languid movement along paths that seemed to suggest unavoidable collisions. I wanted the viewer to wonder what would happen, and when and where.

I could not draw such a four-dimensional work, so I entered a phase of trial and error, learning on a miniature version before working at full size. I widened the inner units to make room for more lead. I cut holes in reinforcing struts to reduce weight. The physical demands of the sculpture determined the size and overall shape of the larger of the two units. Then, for visual harmony, I modified the shapes where structural considerations allowed. Thus, a dialogue between visual desires and technical restrictions dictated the final form.

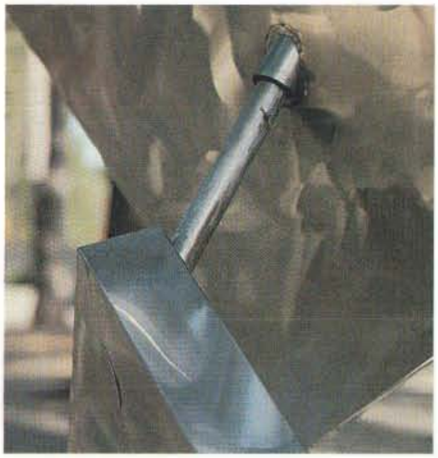
The finished sculpture stood in Berlin for two years and then was bought by Daimler-Benz and permanently installed beside their new administration building in Möhringen, near Stuttgart.

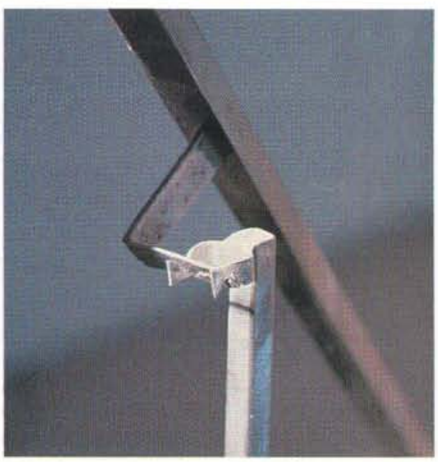
While I was working on the Berlin sculpture, the architect of a new theater in Rotterdam asked if I would be interested in proposing a sculpture for the theater's facade. He sent me blueprints and some information about the site. The theater was surrounded by apartment houses six stories high. My facade was dominated by three 10-meter by 10-meter squares of H-shaped beam that projected five meters from the facade. I used wood sticks to mock up a set of long, slender units linked by shafts set at 45 degrees to one another in order to provide movement across the 15-meter-high wall of the theater, and out from it, filling the space defined by the H beams. Each link was pivoted on its neighbor to create conical movements on three levels.

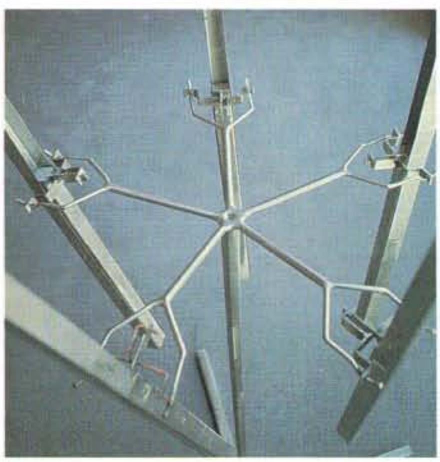
In its final version, the sculpture consists of three columns, each containing four linked segments. The lowest one is a short, sturdy, fixed plinth bolted to the H beam of each steel square. On each support stands a column about six meters high and 35 by 18 centimeters in cross section. The central column rises from its support, whereas those on either side hang down, echoing perhaps the traditional grouping of the Three Graces, two looking forward, the third looking back.

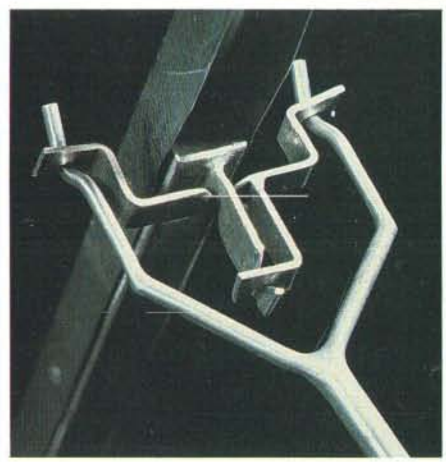
Since completing the Rotterdam work, I have continued to explore the possibilities of multicomponent structures linked by bearings at 90 and at 45 degrees. I have also returned to some simpler kinetic explorations. For example, I mounted a group of four squares so that they would describe an outward (rather than upward) conical movement,

BEARING DESIGN determines how freely the sculptures move and how durable they are. Works designed for the indoors can use finely wrought knife-edge bearings, such as those shown here (*bottom row*). Outdoor works require enclosed ball bearings (*below*), which must be made extremely sensitive to permit motion in even the gentlest breezes. The author works out the motions of his sculptures using scale models (*right*).









like the petals on a flower. This flower-petal motif led me in two directions: toward columns comprising a stack of rotating squares and toward a cluster arrangement of rotating cubes.

I had already tried mounting one square on another, turning on a shaft angled at 45 degrees. I now placed that assembly on another square, also at 45 degrees, then those three on a fourth square moving about a vertical axis. Each square moves along a conical path outward from the one below. The resulting columnar shape appears to break apart and then recover, disintegrating and reintegrating as the squares rotate. One such work was in the Katonah exhibition; a larger version stands beside the new City Hall in Tokyo.

After my experiments with these rotating squares, I began to ponder what would happen if I were to expand the squares into cubes. I found that, on a diagonal corner-to-corner axis, there would be space for them to rotate without collision. But was a cube aerodynamic enough to be set in motion by a light breeze? I made a test model and watched as it slowly answered "yes." So I was able to show four clustered, rotating cubes at Katonah. An enlarged and perfected version of this cube cluster was installed last fall beside the East Wing of the National Gallery in Washington, D.C.

I am now engaged in yet another large-scale project, a hanging sculpture for an eight-story atrium in Berlin, which I will install this coming summer. I know that my artistic explorations will be finite, however. As age slows the hand, the memory and the appetite, I begin to feel an impulse to withdraw and review but also to peek around the next corner to see what still lies ahead. I recall Shakespeare's words for Richard II in prison: "I'll hammer it out./My brain I'll prove the female to my soul,/My soul the father; and these two beget/A generation of still-breeding thoughts,/And these same thoughts people this little world." With a little help from the Muse.

FURTHER READING

CONSTRUCTIVISM: ORIGINS AND EVOLUTION. George Rickey. George Braziller, 1967.

GEORGE RICKEY. Nan Rosenthal. Harry N. Abrams, 1977.

GEORGE RICKEY: ART OF MOVEMENT. Essays by G. Rickey, N. Rosenthal and J. Snyder. Katonah Museum of Art, 1991.

Breaching the Blood-Brain Barrier

Development of a therapy for meningitis has revealed how bacteria penetrate the blood-brain barrier. This knowledge may help physicians treat other disorders of the brain

by Elaine Tuomanen

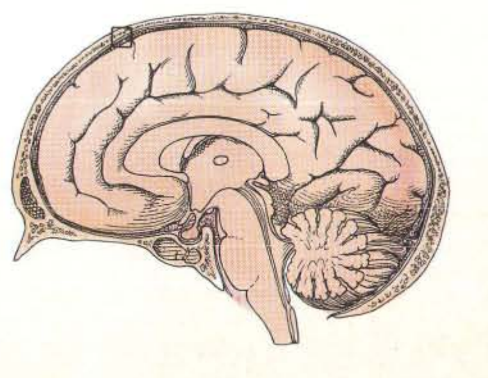
Few frustrations in medicine can match that felt by the neurologist who holds a potent drug within inches of an infection or tumor it cannot reach. The brain alone among the organs of the body remains inaccessible to chemotherapy. Whereas the walls of blood vessels elsewhere in the body are fenestrated, like stone walls with crannies, those of the vessels lining the brain resemble a mortar-sealed brick obstacle. This so-called blood-brain barrier offers free passage to glucose and a few other select substances but fends off intruders, whether friend or foe.

As in all battles, however, spies may infiltrate a line no army could breach. Certain bacteria engage in such deception, slipping through the blood-brain barrier to infect the cerebrospinal fluid. There they induce bacterial meningitis, the most dreaded childhood infection in the U.S. and one of the biggest killers of both adults and children worldwide. In the U.S., it attacks some 50,000 people a year, until very recently killing up to one third of its victims and leaving more than half of the survivors deaf or paralyzed. The figures have been even more dismal for patients whose immune systems have already been compromised by AIDS. My colleagues and I have discovered a great deal about how these bacteria cause meningitis. The work has led to new treatments that drastically improve the survival rate. We may even be on the verge of understanding how the bacterial spies penetrate the blood-brain barrier itself. If we succeed, the tantalizing possibility arises that physicians will soon exploit the spies' secret to smuggle drugs through the blood-brain barrier, for the first time exposing brain tumors and Alzheimer's disease to the entire armamentarium of modern medicine.

ELAINE TUOMANEN heads the laboratory of molecular infectious diseases at the Rockefeller University and is an attending physician at the hospital affiliated with it. She studied at McGill University, earning a B.Sc. in biochemistry in 1973 and an M.D. in 1977. (That year she also was elected to Canada's national basketball team for women.) She trained as a pediatrician at Montreal Children's Hospital and held a fellowship in infectious diseases at the University of Virginia. In 1982 she moved to Rockefeller, where she was Bristol Fellow of Infectious Diseases for one year and Parker B. Francis Fellow of Pulmonary Research for another three years. She has been a member of the university's faculty since 1984.

Three bacteria can cause meningitis. *Hemophilus influenzae* (which has nothing to do with influenza proper) was the most common until 1990, when the widespread administration of a new vaccine promised to stamp it out. *Neisseria meningitidis*, or meningococcus, continues to cause epidemics in Africa, where the vaccine is poorly distributed. Most treacherous of all is *Streptococcus pneumoniae*, or pneumococcus, which also causes pneumonia and infections of the middle ear. Pneumococcal meningitis killed about 30 percent of the children it infected, no matter how intensively they were treated, until two years ago, when my colleagues and I found a way to reduce mortality to well under 5 percent. The three forms of bacterial meningitis all take a single course. The patient develops the fever, irritability and drowsiness characteristic of viral influenza; these are followed by brain inflammation that can cause seizures, coma and, in many cases, death. The physician therefore looks for the subtle signs of meningitis—above all, a feeling of stiffness in the neck. The mere suspicion of the disease signals a medical emergency, compelling immediate infusion of antibiotics into the patient's veins. When antibiotics are given in very high doses, just enough will usually leak through the blood-brain barrier to kill most bacteria. But the physician must mix and administer a veritable cocktail of drugs; the luxury of waiting a day to identify the invader and choose a specific antibiotic does not exist.

For the next 24 hours, two questions vie for attention: Is the antibiotic killing the bacteria? and Will the patient live? The questions, we have found, are tragically linked. Drugs of the penicillin family utterly eradicate the three agents of meningitis, yet one out of three children still dies. This is the problem that frustrated all of us in the Cooperative Meningitis Study Group, an informal association that meets every other year. If the infection is cured, why does the patient die?



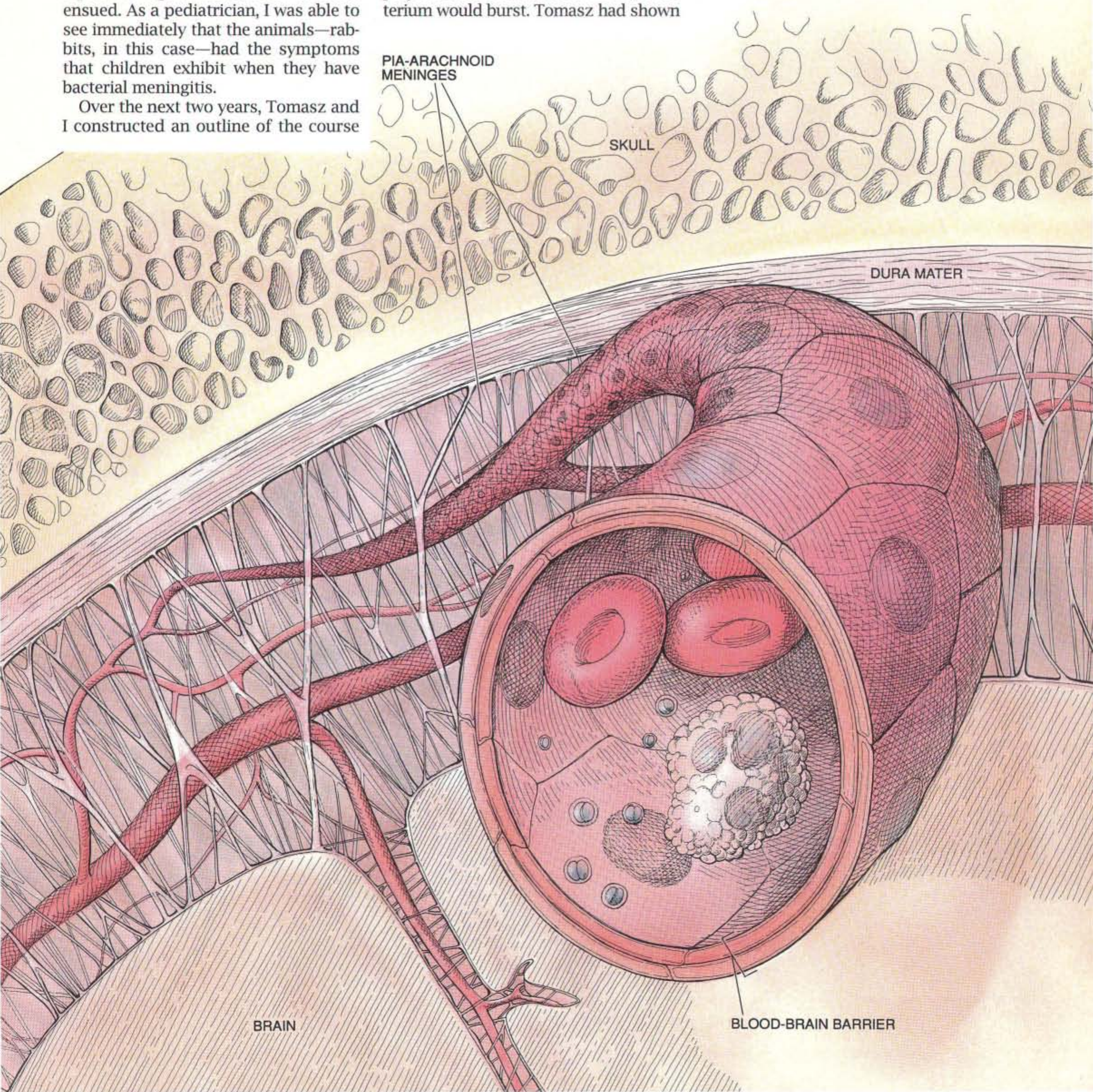
BLOOD-BRAIN BARRIER, a tightly tiled lining of the brain's vascular system, excludes most substances. Meningitis occurs when bacteria in the bloodstream slip past the usually impenetrable barrier and multiply in the cerebrospinal fluid.

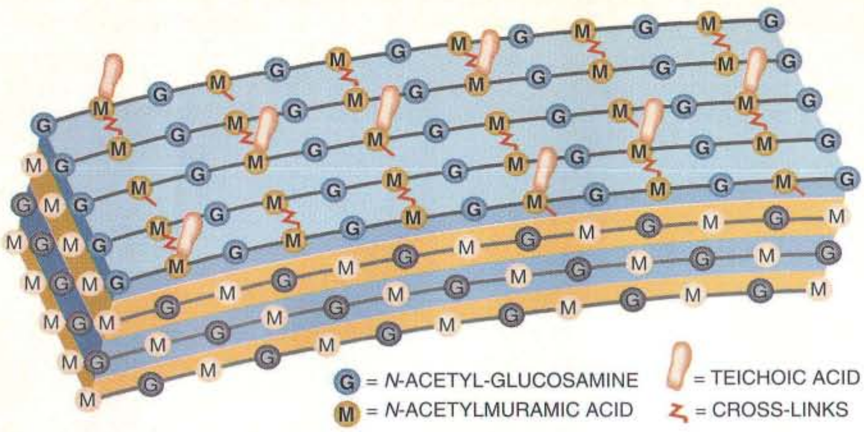
I remember the day in the summer of 1984 when I began to understand how to get at the answer. My colleague Alexander Tomasz called me from Switzerland, where he was spending a sabbatical working at the Ciba-Geigy Research Laboratory in Basel. He had just been testing in an animal model of meningitis how antibiotics kill bacteria. Accidentally, he had challenged the animals with killed pneumococci, and yet they still got sick. Dead bacteria seemed as harmful as living ones. How could this be?

That week I flew to Basel to help him repeat the experiment. The same results ensued. As a pediatrician, I was able to see immediately that the animals—rabbits, in this case—had the symptoms that children exhibit when they have bacterial meningitis.

Over the next two years, Tomasz and I constructed an outline of the course of meningitis, using a rabbit model of the disease that Ralph G. Dacey, Jr., and Merle A. Sande had developed in the early 1980s at the University of Virginia. Our collaboration showed that bacterial corpses do not kill directly; instead they stir the body into a self-destructive frenzy that leads to the symptoms and consequences of the disease.

The fragments of the bacterial corpse that roil the immune system come from the cell wall, a rigid exoskeleton made of chain-linked polysaccharides. Without the wall, a bacterium would burst. Tomasz had shown that most bacteria, including the pneumococcus, disintegrate as they die in a fascinating process that involves the dissolution of the cell wall. Penicillin takes advantage of this suicidal tendency by triggering lysis, a process that involves a bacteria's own enzymes, which normally nip and tuck the wall into a perfect fit. Penicillin destroys the controls that restrain the enzymes, so the enzymes tear the cell wall to pieces. Ex-





BACTERIAL CELL WALL consists of a single cross-linked molecule, a portion of which appears above in simplified form. Penicillin shatters the wall into millions of its 20 or so building blocks, each of which contributes to a symptom of infection.

actly how this happens remains a mystery almost 50 years after penicillin was discovered. Now, however, we had a clue that this explosive death has a dark side that injures the body even as it ends the infection.

Our experimental model showed that when the bacteria multiply in the cerebrospinal fluid the animal at first shows no symptoms. When the concentration of bacteria reaches 100,000 per milliliter of fluid, however, it is as if a spell is broken. Fever and inflammation break out, as the body's defenses suddenly deploy hormonelike substances called cytokines (interleukin-1 and tumor necrosis factor). The cytokines issue from the brain's side of the blood-brain barrier and force the barrier to let in the white blood cells. Dead bacteria induce the same course of events and at the same concentration: 100,000 per milliliter. In an untreated patient, then, the bacterial debris as well as living bacteria encourage the cycle of inflammation and its consequences.

How could killed bacteria cause the body to do itself such harm? Could the immune system be mistaking the dead for the living? To find out, we challenged the animals with just 1,000 living pneumococci and immediately treated them with penicillin. Then we waited to see whether the exploding bacteria would induce white blood cells to storm the brain. Thirty minutes later the cerebrospinal fluid was indeed swarming with white blood cells.

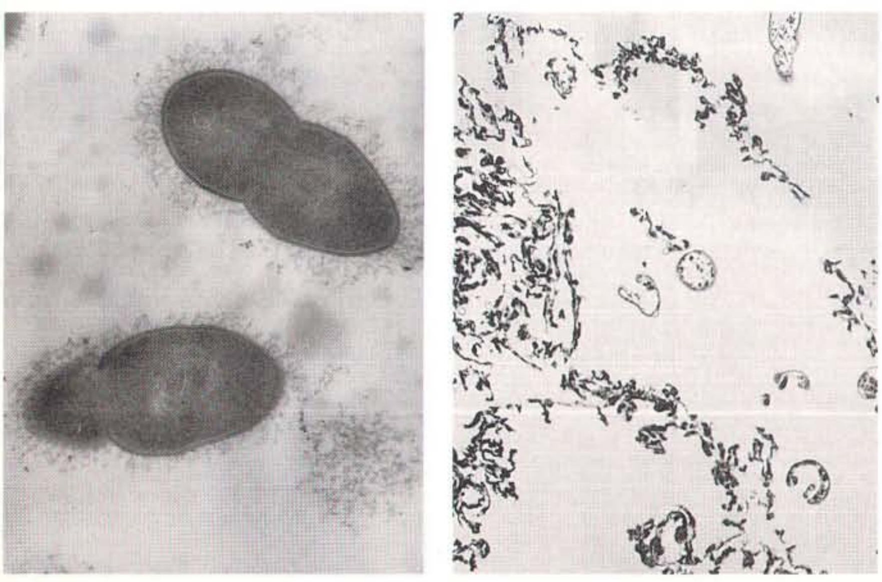
Soon after these experiments, Martin G. Täuber, Jay H. Tureen and Sande completed similar studies at San Francisco General Hospital. This work showed that not only inflammation but also brain swelling and intracranial pressure increased to dangerous levels in the first few hours of antibiotic therapy. Such violent inflammation defaces the cellular architecture of the brain and supporting tissues.

Our hunch had been confirmed. When a bacterium shatters, body defenses mistake the shrapnel for a burst of bacterial growth and respond accordingly. The bits of cell wall activate the body's defenses by setting off the cytokine alarms and the clotting initiators. These events prepare a platform on the vessel wall to which white blood cells stick, giving them enough purchase to squeeze themselves through the blood-brain barrier. The white blood cells, now ensconced in the brain, enhance the production of cytokine alarms, accelerating the disruption of the barrier. More white cells migrate into the brain, further exacerbating inflammation, swelling and the release of toxic by-products of the immune reaction.

In short, the animal model had taught us an unexpected lesson: antibiotic therapy makes meningitis worse before it makes it better. Armed with this critical information, clinicians working with George H. McCracken, Jr., of the University of Texas Southwestern Medical Center at Dallas were able to show that the same deterioration occurred in children treated for meningitis. All patients would die without antibiotics, but clearly some also do not survive the cure. The remainder lose critical nerve cells, and many are left with deficits in hearing, movement or learning.

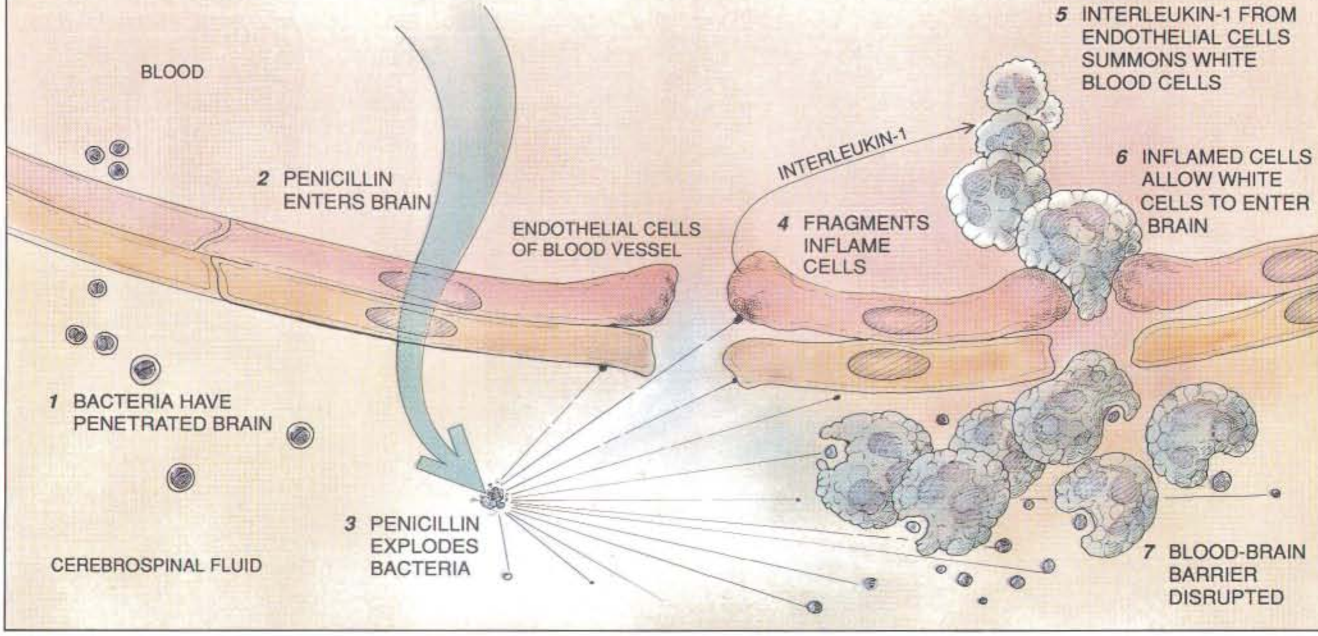
To save the patient from the cure, the physician must find a way to calm the errant host defenses while the antibiotics do their killing. To determine which defense systems must be suppressed and which must not be, one needs a detailed understanding of all the events in the genesis of the disease. Although such information is being developed by the Cooperative Meningitis Study Group, final results will take years. In the meantime, some of us are looking for a provisional solution.

The most powerful anti-inflammatory drug, the steroid dexamethasone, virtually shuts down the inflammatory process. We believed, therefore, that it would almost certainly prevent the host defenses from stampeding during the critical early hours of antibiotic therapy. Yet although years of clinical experience had proved the drug effective in inflammatory settings, such as arthritis, its record was inconsistent in infectious settings. Moreover, some physicians continued to resist the idea of shutting down the body's defenses at



BURSTING BACTERIA: intact pneumococci have a fuzzy capsular layer on their cell walls (*left*). Penicillin causes the bacteria to destroy their own walls and burst, littering the microscope's field with cell-wall debris (*right*).

The Pathology of Meningitis



the very peak of infection. For these reasons, we were compelled to try the drug in the experimental model before moving to children.

We elected to test steroids and their equally potent but more selective sisters, the nonsteroidal anti-inflammatory agents. Luckily, the answer was decisive and quick. Working with Oto Zak of Ciba-Geigy Research Laboratories, we showed that the administration of both kinds of anti-inflammatory drugs not only blocked the inflammatory response to bacterial death but also enhanced animal survival. For some of the nonsteroidal treatments, mortality decreased nearly to zero [*see right illustration on next page*].

Unfortunately, the Food and Drug Administration has not approved these nonsteroidal anti-inflammatory drugs (such as oxindanac) for infectious diseases in children. Still, steroids could be used, and in 1987 McCracken's group in Dallas started a trial combining steroids with antibiotics. The steroids decreased the incidence of hearing loss from 15 to 3 percent and shortened the course of the fever from five days to 1.6. These results were at first received with skepticism. Since then, many centers throughout the world have confirmed these figures, and now it is recognized that the concepts learned from the experimental model have held true.

In 1990 the American Academy of Pediatrics adopted the use of steroids together with antibiotics as the recommended therapy for childhood meningitis. W. Michael Scheld and his colleagues at the University of Virginia are now striving to demonstrate the benefit of steroids in meningitis in a trial for adults. But steroids have significant side effects, such as gastrointestinal bleeding. Such bleeding might prove serious in a child already weakened by infection. I believe other therapies can do much better.

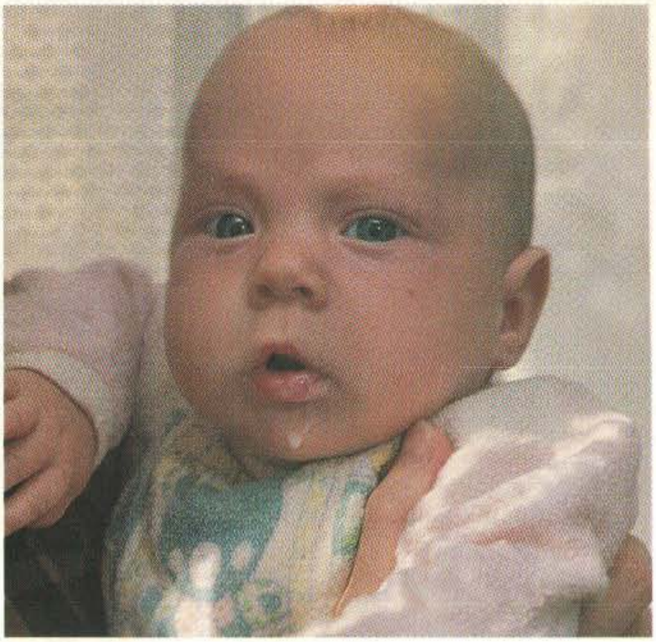
The basic problem is the brutal way in which most antibiotics kill bacteria. If the bacteria did not fall apart, there would be no problem. One wishes, therefore, for a more refined breed of executioner. A new family of antibiotics derived from penicillin, called the penem family, comes close to this ideal. Although these antibiotics also break the cell wall, the fragments are less noxious than those elicited by penicillin. Another strategy is to neutralize cell-wall fragments. The sure and specific way of doing this is with antibodies tailor-made to mask the fragments. This masking should prevent the host defenses from recognizing the fragments as foreign bodies. But here again we return to the problem of the blood-brain barrier, which bars the passage of antibodies.

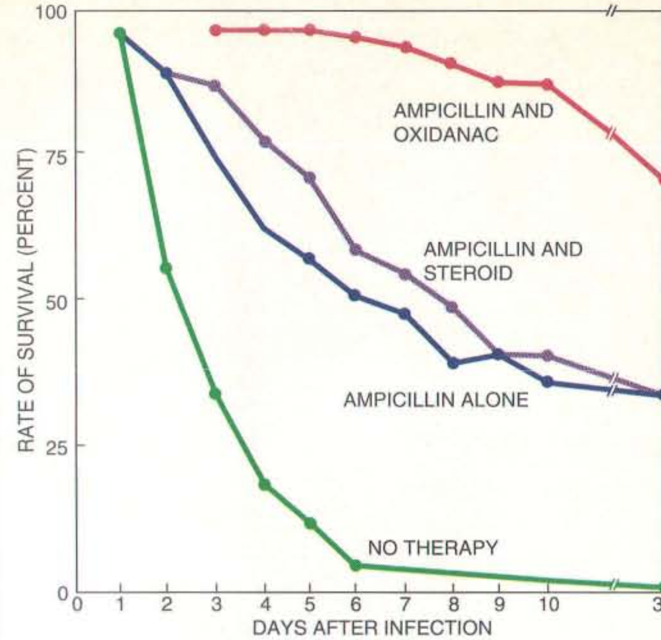
If we want to smuggle antibodies into the brain, we must find a way to pry open the barrier. Yet a breaching of the wall would disturb the nerve cells, which inhabit a milieu completely different from that of the blood serum. Such disturbance might contribute to the injury seen in meningitis. Would it not be better, then, to find an antibody that could keep white blood cells from entering the brain in the first place?

To achieve this effect, I tried using an antibody, called anti-CD18, that is known to prevent white blood cells from sticking to the vessel wall. Thus treated, the cells should be prevented from crossing over into the brain tissues for a few hours—enough time for the fireworks of bacterial death to subside. To test the hypothesis, I injected anti-CD18 into the animal model, then administered penicillin.

The events of the next few hours are among the most impressive in my memory. The infected animals that received the antibody together with the antibiotics rapidly became alert and resumed normal activity as if they had not been ill; the survival rate was 100 percent. The benefit of preventing disruption of the blood-brain barrier by white blood cells was tremendous. It is anticipated that this antibody will be available for clinical trials in humans in several months.

Meningitis has been bowed but not beaten once and for all. To do that, one would like to reduce the course of meningitis to a sequence of events, identify the deleterious ones and devise a specific counter to each. We know that bioactive pieces of the burst bacterium, mainly from the cell wall, encode the information necessary to provoke all the signs and symptoms of disease. We therefore hypothesize that each piece helps to cause one or more of the symptoms. Because each bacterial species has its own





AUTHOR'S TWO-STEP CURE freezes the immune system, then kills the bacteria. This child recovered. Studies suggest that average mortality has fallen from 30 to less than 5 percent.

EVEN LOWER MORTALITY should result when nonsteroidal anti-inflammatory agents are approved for use in children. Animal model shows optimum results for oxindanac.

cell-wall structure and its own set of enzymes to disrupt it, bacterial infections will differ in a predictable way. One needs to study the wall-building machinery of each species.

We benefit from a technique devised in Germany to define the library of wall subcomponents encasing each of the major human pathogens. The wall of each invader contains about 20 of these building blocks, which our laboratory and others are striving to characterize. Recent work has shown that, true to prediction, different pieces have different biological activities.

This rule appears to apply to infections in other parts of the body as well. We have shown that wall pieces cause inflammation in the lung and ear. Raoul S. Rosenthal of the Indiana University School of Medicine and William E. Goldman of Washington University have shown that one piece, common to many bacterial cell walls, can cause arthritis and kill cells lining the airways. James M. Krueger of the University of Tennessee Center for Health Sciences has determined that another piece induces sleep and probably the coma seen in advanced infection. Perhaps further study will show whether small doses can serve as a safe sleeping pill.

Our discoveries of the effects of cell-wall pieces on many body functions have led us to collaborate with Alkermes, Inc., in Cambridge, Mass., to try to turn the tables in this spy story. We plan to use the abilities of the invader to our own advantage. And if we can wring drugs from the dead body of the pneumococcus, what cornucopia might we find in other bacterial species?

I have discovered one particularly interesting possibility in *Bordetella pertussis,* the agent of whooping cough. It fools the body by producing chemical signals that interfere with the ability of patrolling white blood cells to recognize the address at their destination. Physicians might use the bioactive part of this bacterium to devise a highly selective anti-inflammatory agent to treat inflammation in the brain, joint, eye and other sites where the surrounding tissue must be protected from the immune system as well as from invaders.

Such speculation leads us to seek the most exciting goal of all: the bacterial signal that enables the organism to penetrate the blood-brain barrier and cause illness. Tomasz and I look for the key to the blood-brain barrier by cutting up the cell wall of the pneumococcus with the very enzymes the bacterium uses to undergo lysis. Then we pass the pieces through a high-pressure liquid chromatograph to separate them according to their chemical composition. Next we inject each kind of piece into a rabbit. Finally, we check to see if we have opened any doors into the brain by injecting the animal with tracer molecules that normally cannot enter the brain. If the tracer finds its way into the organ, we know we have safely opened the door.

As this article goes to press, we believe we have found a candidate substance: a glycopeptide, a protein linked with a sugar. This may be only one of several keys to the blood-brain barrier. We are looking for others.

Natural selection has furnished the bacterial cell wall with many passwords, each of which triggers some normal human process. These passwords could constitute a treasure of therapeutic devices. One might compare the bacterial cell wall with the natural toxins of plants and animals—curare, for example—that medicine has harnessed for the benefit of humanity. We have learned to control the riot of host defenses that dying bacteria incite. We have the suspected agitators under interrogation. The silent spy that infiltrates the brain's defenses cannot long elude us.

FURTHER READING

PARTNER DRUGS: A NEW OUTLOOK FOR BACTERIAL MENINGITIS. E. Tuomanen in *Annals of Internal Medicine,* Vol. 109, No. 9, pages 690–692; November 1, 1988.

MOLECULAR PATHOPHYSIOLOGY OF BACTERIAL MENINGITIS: CURRENT CONCEPTS AND THERAPEUTIC IMPLICATIONS. X. Saez-Llorens, O. Ramilo, M. M. Mustafa, J. Mertsola and G. H. McCracken, Jr., in *Journal of Pediatrics,* Vol. 116, No. 5, pages 671–684; May 1990.

BACTERIAL MENINGITIS: PATHOGENESIS, PATHOPHYSIOLOGY AND PROGRESS. V. Quagliarello and W. M. Scheld in *New England Journal of Medicine,* Vol. 327, pages 864–872; September 17, 1992.

60% will lose...

but 40% will win up to $4 million!

GOVERNMENT CONTROLLED GERMANY

NEW!

Our TOP PRIZES now: 2x $4,284,000!

4 drawings a month - every month jackpots between $714,000 and $4,284,000!

Almost every second number wins ...
in the famous German „Staatliche Nordwestdeutsche Klassenlotterie". The 90th lottery starts on March 26, 1993 with the fantastic prize money of $433,029,933! That is something you surely can rely on, as it is all guaranteed by the German Government! So don't miss your big win this time. 26 drawings with 2x $4.284 million, 2x $3.57 million, 2x $2.856 million, 2x $2.142 million, 1x $1.428 million and 27x $714,000. Plus thousands of prizes worth up to $178.500! Order your ticket(s) on the coupon below. You will receive them within days, as well as the relevant rules and regulations.

Your winning chances: **Playing 1 ticket number = 40,1%**
Playing the 5/8-Combi = 92,8%

Don't miss your big win this time ...
...enter the 90th North-West German State Lottery. Send payment with your order or pay after receipt of your statement of account either by cheque by travellers' cheque, bank transfer, cash via registered air mail (at your own risk) or by credit card (Handling charge for payment by credit card: 6%.). And remember... **The bigger your stake, the bigger your win!** Prices, including charges for airmail postage and the Official Winning Lists, are to be understood for the whole lottery – 26 drawings, one drawing per week.

Please fill in the number of tickets you want to order.

		DM	US $*	£*
	1/1 ticket	864,00	617.50	377.00
	1/2 ticket	444,00	317.50	193.50
	1/4 ticket	234,00	167.50	102.00
	5/8 combi	549,00	392.50	239.00

*Rate of exchange may vary. Amounts are calculated: $1.00 = DM 1,40/ £1.00 = DM 2,30.

Correspondence only possible in German or English! 0804

Answer in German ☐ in English ☐ Mail ticket to: Mr.☐ Mrs.☐ **Please print in block letters.**

First name: ______

Surname: ______

Street/No.: ______ P.O.Box: ______

City, Country, Post Code: ______

☐ Cheque included

☐ Charge to my credit card: Visa ☐ American Express ☐ Diners ☐ Euro/MasterCard ☐

Expiration Date: Month ______ Year ______ Signature X ______

Valid only where legal.

SERVICE WORLD-WIDE HERZOG Lotterie-Einnahme Hans Herzog

State licensed Lottery Agent HANS HERZOG
Grützmühlenweg 46
W-2000 Hamburg 63
Germany

Redeeming Charles Babbage's Mechanical Computer

A successful effort to build a working, three-ton Babbage calculating engine suggests that history has misjudged the pioneer of automatic computing

by Doron D. Swade

Charles Babbage is celebrated as the great ancestral figure in the history of computing. The designs for his vast mechanical calculators rank among the most startling intellectual achievements of the 19th century. Yet Babbage failed in his efforts to realize those plans in physical form. Histories of computing routinely assert that Babbage faltered primarily because the demands of his devices lay beyond the capabilities of Victorian mechanical engineering. Curiously, no contemporary evidence supports that view.

In 1985 my colleagues and I at the Science Museum in London set out to resolve or at least illuminate the question by building a full-size Babbage computing engine based on his original designs. Our endeavor finally bore fruit in November 1991, a month before the bicentenary of Babbage's birth. At that time, the device—known as Difference Engine No. 2—flawlessly performed its first major calculation. The success of our undertaking affirmed that Babbage's failures were ones of practical accomplishment, not of design.

Those failures have become inextricably associated with his creative genius. Babbage, proud and principled, was famed for the vigor and sarcasm of his public denunciations of the scientific establishment. The demise of his engine project added a sense of injustice, bitterness and even despair to his celebrated diatribes. Since then, he has acquired an image of testiness and eccentricity; the first biography of Babbage, written by Maboth Moseley and published in 1964, was titled *Irascible Genius: A Life of Charles Babbage, Inventor.* Our work at the Science Museum emphasizes a distinctly different side of Babbage: a meticulous inventor whose designs were hugely ambitious but well within the realm of possibility.

CHARLES BABBAGE sat for this daguerreotype around 1847, the year he began work on Difference Engine No. 2.

Babbage's desire to mechanize calculation arose from the exasperation he felt at the inaccuracies in printed mathematical tables. Scientists, bankers, actuaries, navigators, engineers and the like relied on such tables to perform calculations requiring accuracy to more than a few figures. But the production of tables was tedious and prone to error at each stage of preparation, from calculation to transcription to typesetting. Dionysius Lardner, a well-known popularizer of science, wrote in 1834 that a random selection of 40 volumes of mathematical tables incorporated 3,700 acknowledged errata, some of which themselves contained errors.

Babbage was both a connoisseur of tables and a fastidious analyst of tabular errors. He traced clusters of errors common to different editions of tables and deduced where pieces of loose type had been incorrectly replaced after falling out. On one occasion, he collaborated with John Herschel, the renowned British astronomer, to check two independently prepared sets of calculations for astronomical tables; the two men were dismayed by the numerous discrepancies. "I wish to God these calculations had been executed by steam!" Babbage exclaimed in 1821.

Mechanical computers should, Babbage thought, offer a means to eliminate at a stroke all the sources of mistakes in mathematical tables. He envisioned a machine that not only would calculate flawlessly but would eradicate transcription and typesetting errors by automatically impressing the results of its calculations onto papier-mâché strips or plates of soft metal. A printed record could then be generated directly from those plates, thereby eliminating every opportunity for the genesis of errors.

In 1822 Babbage built an experimental model intended to carry him toward his goal. He called his mechanical calculator a "difference engine" because it is based on a mathematical principle known as the method of finite differences. The method permits one to determine successive values of polynomial functions using only addition [*see box on page 66*]. Multiplication and di-

DORON D. SWADE is both an electronics engineer and a historian of computing. He has been senior curator of the computing and control section of the Science Museum in London since 1985 and has published articles on curatorship and on the history of computing. He has recently written two books: *Charles Babbage and His Calculating Engines,* which accompanies the Babbage exhibition that Swade curated, and, in collaboration with Jon Palfreman, *The Dream Machine: Exploring the Computer Age,* a companion to the television series of the same name. Swade led the project to construct a full-scale Babbage calculating engine.

vision, which are far more difficult to mechanize, are not necessary. Because the value of the function at each step is calculated based on its predecessor, a correct final result imparts a high degree of confidence that all previous values are also correct.

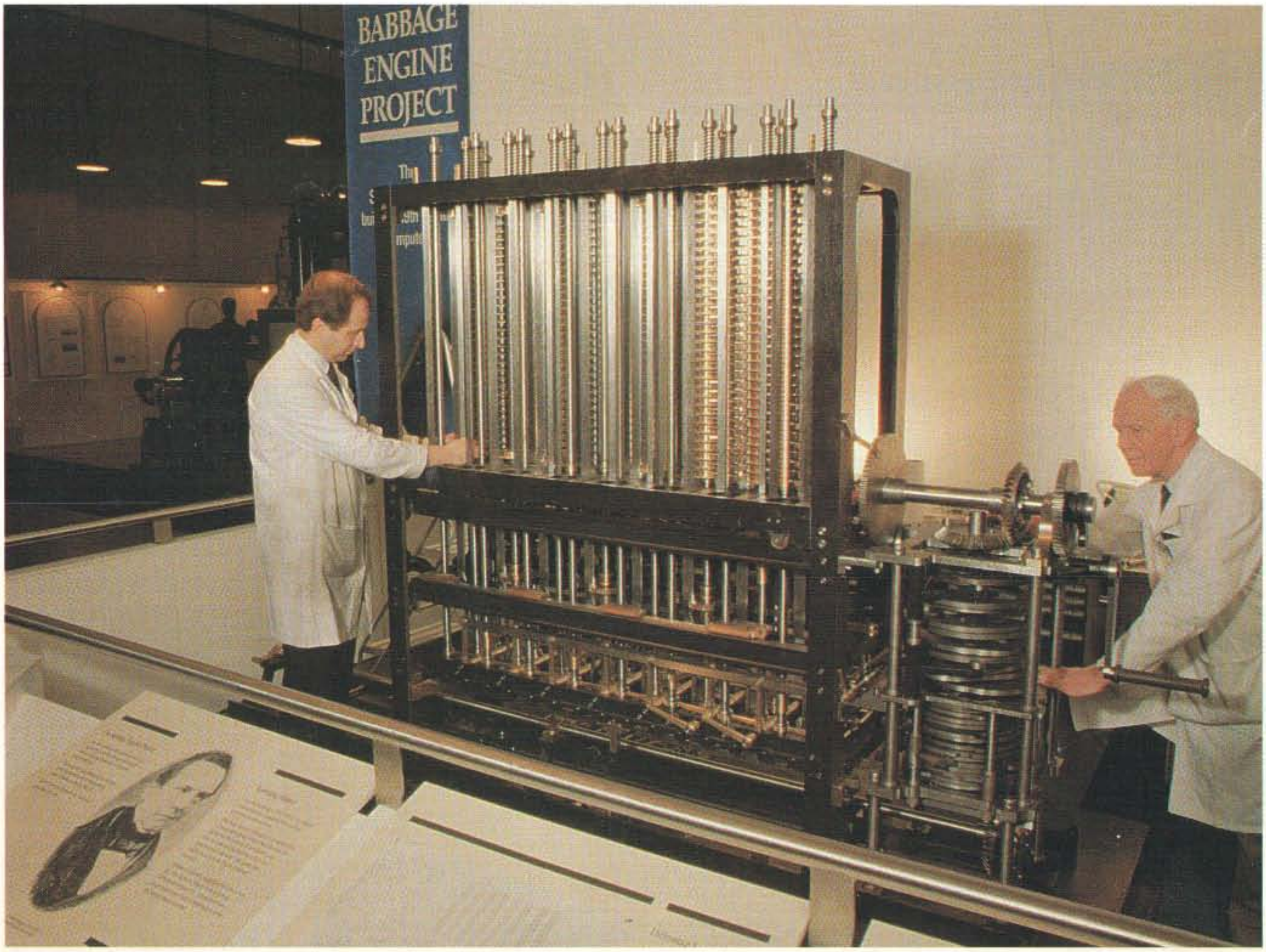
For economy of design, Babbage's difference engines use the decimal number system rather than the binary system common to modern electronic computers. Each digit in a multidigit number is represented by a toothed gear wheel, or figure wheel, engraved with decimal numerals. The value of each digit is represented by the angular rotation of the associated figure wheel. The engine's control mechanism ensures that only whole-number values, represented by discrete positions of the figure wheels, are valid. Babbage boasted that his machines would produce the correct result or would jam but that they would never deceive.

Babbage's most ambitious venture to construct a full-scale calculating device was devoted to the ill-fated Difference Engine No. 1. His efforts foundered in 1833 after a decade of design, development and component manufacture, not to mention vast expense. The project collapsed after a dispute between Babbage and his chief engineer, Joseph Clement, over payment for relocating the machining works. Outwardly at least, technology did not feature in the disagreement. The question that has remained tantalizingly unresolved is whether the circumstances surrounding the collapse of the project concealed the technical or logical impossibility of Babbage's schemes.

Difference Engine No. 1 consists of a basic adding element, repeated many times over in an arrangement that embodies the method of differences. The size and complexity of the engine are monumental: the design calls for roughly 25,000 parts; the assembled machine would measure eight feet high, seven feet long and three feet deep; and it would weigh several tons. The project, which was funded by the British government, was also enormously expensive. When Clement's last bill was paid in 1834, the cost totaled £17,470. For comparison, the steam locomotive John Bull, built in 1831, cost all of £784.

Clement completed about 12,000 of the 25,000 parts required for Difference Engine No. 1, most of which were later melted down as scrap. The British government finally withdrew from the project in 1842, partly on the advice of George Biddell Airy, Astronomer Royal, who pronounced Babbage's engine "worthless." The failure to complete the difference engine was the central trauma in Babbage's scientific life; it is a topic he returns to repeatedly in his writings as though unable to reconcile himself to the dismal outcome.

The years of work on Difference Engine No. 1 did produce one noteworthy, tangible result. In 1832 Clement assembled a small section of the engine, con-



DIFFERENCE ENGINE NO. 2 was constructed in public view at the Science Museum in London. Here the two engineers who built it, Barrie Holloway (*left*) and Reg Crick (*right*), perform some essential adjustments. Babbage also designed a printing mechanism for the difference engine, but because of limited time and money, the printer has not yet been built.

sisting of about 2,000 parts, as a demonstration piece. This finished part of the unfinished engine is one of the finest examples of precision engineering of the time and works impeccably to this day.

The demonstration piece is the first known automatic calculator. Unlike the desktop calculators of the time, the engine, once set up, did not rely on informed human intervention. Thus, an operator could achieve accurate results without any understanding of the logical or mechanical principles involved. The opportunity to speculate about machine intelligence was not lost on Babbage and his contemporaries. Harry Wilmot Buxton, a younger colleague with whom Babbage entrusted many of his papers, wrote that "the wondrous pulp and fibre of the brain had been substituted by brass and iron; he [Babbage] had taught wheelwork to think."

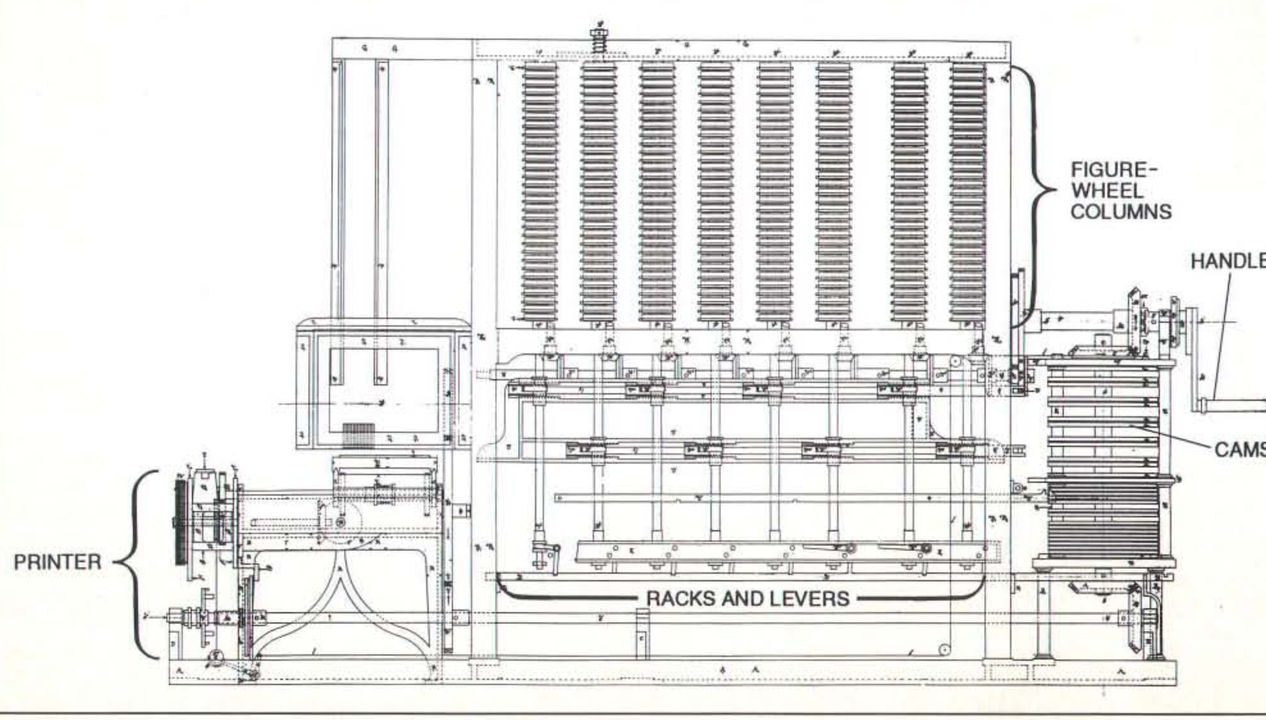
Despite its impressive capabilities, the difference engine could perform only one fixed task. Babbage's reputation as a computer pioneer largely rests on another, more sophisticated device—the Analytical Engine, conceived by 1834. He intended the Analytical Engine as a general-purpose programmable computing machine, whose features are startlingly similar to those of modern electronic computers. It had a basic repertoire of operations (addition, subtraction, multiplication and division) that it could execute in any sequence. The internal architecture of the machine featured a separate "store" and "mill," equivalent to the memory and processor in a modern computer. The separation of store and mill has been a dominant design feature of electronic computers since the mid-1940s.

The Analytical Engine could be programmed by using punched cards, a technique previously used in the Jacquard loom to control patterns of woven thread. The Analytical Engine could take alternative courses of action depending on the result of a calculation, enabling it to perform complex functions. Babbage intended the machine to be able to handle up to 50-digit input numbers and 100-digit results; the output could be printed, punched or plotted.

Although historians customarily refer to the Analytical Engine as if it were a physical thing, it is actually a series of unbuilt designs that Babbage refined at intervals from 1834 until his death in 1871. Demoralized by the fate of Difference Engine No. 1, he made no serious attempt to construct a full-scale Analytical Engine. A small experimental part of the mill that was still incomplete at the time of his death, along with

How Babbage's Difference Engines Work

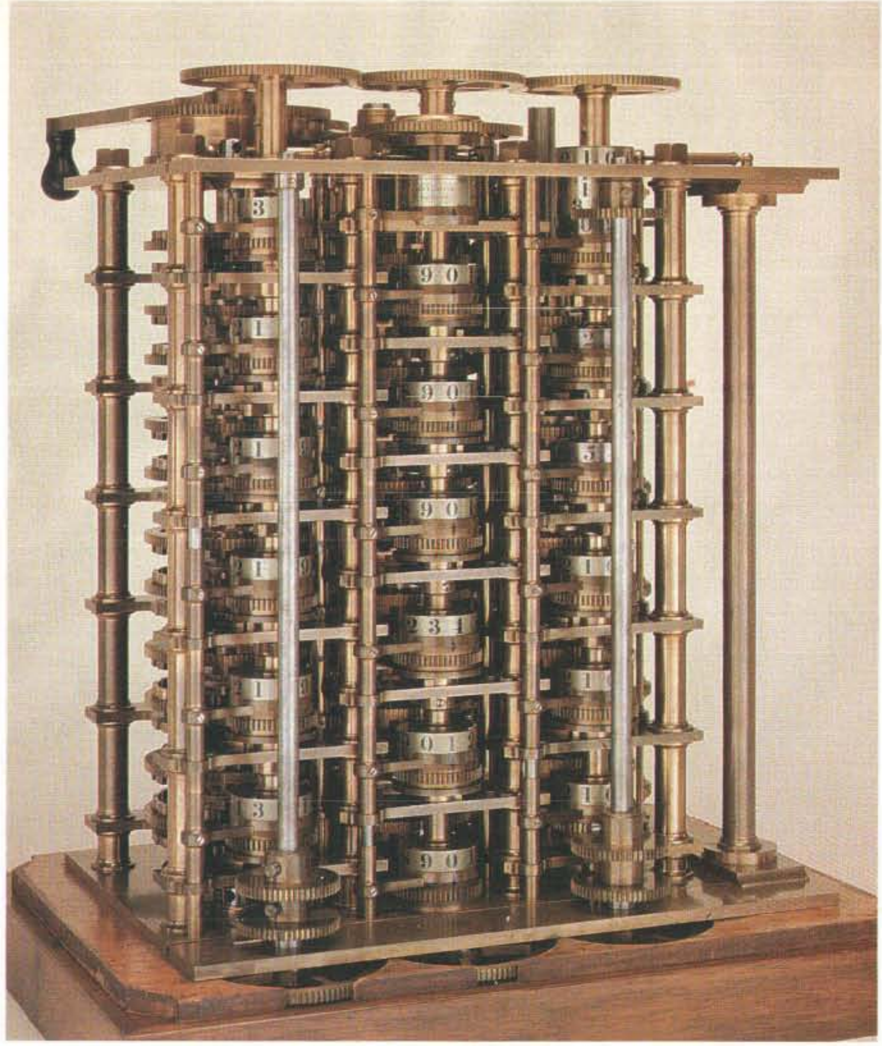
Shown below is one of Babbage's 20 main drawings of Difference Engine No. 2, which he drafted in 1847. The machine is operated by means of the handle on the right. Turning the handle rotates a vertical stack of 14 pairs of cams that determine the action and timing of the calculating cycle. Numbers are stored and operated on in eight vertical columns, each of which contains 31 engraved figure wheels. The least significant digit of a number is stored at the bottom of the column, the most significant digit at the top. The initial values for a calculation are entered by unlocking the figure wheels and rotating each one by hand to the appropriate decimal value. Below the figure-wheel columns are a set of racks and levers that, when activated by links from the cams, lift, lower and turn the vertical axes, thereby carrying out the addition of differences. Difference Engine No. 2 does not add numbers in sequence from right to left, as one might expect. Instead values from odd-numbered columns are added to even-numbered columns during the first half-cycle; even-numbered columns are then added to odd-numbered columns during the second half-cycle. This technique significantly reduces the time required for a calculation. A similar approach, known as pipelining, is used in modern electronic computers. The printing assembly, located at the left, is directly coupled to the last column of figure wheels, which bear the final result of the calculation. Each turn of the handle produces one 30-digit value in the table of differences and automatically prepares the machine to generate the next number.



another fragment later built by Babbage's son, Henry Prevost Babbage, are the only significant remains of his grand designs.

Work on the Analytical Engine forced Babbage to think about how to develop mechanisms capable of automatic multiplication and division, all regulated by a complex control system. The solutions to those problems inspired him to design a simpler and more elegant difference engine, Difference Engine No. 2. Although the machine calculates to a precision of 31 figures, 10 digits more than Babbage envisaged for Difference Engine No. 1, it contains only one third as many parts. Babbage drew up detailed plans for the second machine between 1847 and 1849 and offered them to the government in 1852 but received no encouragement. So things stood for nearly a century and a half.

During several visits to London beginning in 1979, Allan G. Bromley of the University of Sydney in Australia examined Babbage's drawings and notebooks in the Science Museum Library and became convinced that Difference Engine No. 2 could be built and would work. I had independently read of Babbage's hapless fate and become deeply puzzled as to why no one had tried to resolve the issue of Babbage's failures by actually building his engine.



WORKING PART of Difference Engine No. 1, assembled by Joseph Clement in 1832, is the first known automatic calculator. Its flawless operation strongly supports Babbage's conviction that building a full-sized engine was a practical prospect.

In 1985, shortly after my appointment as curator of computing, Bromley appeared at the Science Museum carrying a two-page proposal to do just that. He suggested that the museum attempt to complete the machine by 1991, the bicentenary of Babbage's birth. Bromley's proposal marked the start of a six-year project that became something of a personal crusade for me. The saga of our effort to construct the difference engine is one worthy of Babbage himself. We embarked on a complex engineering project that took us into unknown technical territory and confronted us with mechanical conundrums, funding crises and the intrigues inherent in any major venture.

Difference Engine No. 2 was clearly the engine of choice for the project. The associated set of drawings is intact, whereas those for Difference Engine No. 1 show regrettable gaps. Difference Engine No. 2 is also a more economic design. Cost and time constraints argued in favor of ignoring the printer and concentrating on the rest of the engine. The printer is composed of about 4,000 parts and would be a sizable engineering project in its own right.

The documentation for Difference Engine No. 2 consists of 20 main design drawings and several tracings. As we pored over those drawings, my colleagues and I discovered several flaws in the plans, in addition to those identified by Bromley. One major assembly appears to be redundant. Other mechanisms are missing from the design. For example, the initial values needed to begin a calculation are entered by unlocking the columns and manually rotating each of the freed figure wheels to the appropriate positions. Babbage omitted a means of locking the columns after they were set, so the setting-up procedure was self-corrupting.

The most serious design lapse concerned the carriage mechanism. This crucial component ensures that if, in the course of an addition, the value on a figure wheel exceeds 10, then the next higher figure wheel (indicating numbers 10 times larger) advances one digit. The most extreme test of the carriage mechanism occurs when a 1 is added to a row of 9's. Babbage solved the carry problem in an exquisitely innovative manner. During the first part of the calculating cycle, the engine performs a 31-digit addition without carrying the 10's, but every figure wheel that exceeds 10 sets a spring-loaded warning device. In the second part of the cycle, each armed warning device allows a rotating arm to advance the next higher figure wheel by one position.

Unfortunately, the configuration of the carry mechanism shown in Babbage's design drawings is unworkable. The direction of rotation of the figure wheels is incorrect, and the warning-and-carry mechanism could not function as drawn. The source of these shortcomings stimulated considerable speculation. We considered the possibility that errors were introduced deliberately as security against industrial espionage. More likely, some flaws were design oversights, and others were inevitable drafting and layout errors.

None of the design problems we found in Difference Engine No. 2 compromised its overall logic or operational principles, and we managed to devise solutions for all. Unnecessary mechanisms were omitted. The missing locking assemblies for the figure wheels were devised and, where necessary, their motions derived from those of neighboring pieces. Bromley solved the carry-mechanism problem by mirror-reversing the incorrectly drawn parts and altering their orientation. The introduction of a four-to-one reduction gear in the drive allayed skepticism about whether the massive Difference Engine No. 2 could be driven by hand. This change made the drive handle four times easier to turn but caused the engine to run four times slower.

Implementing the solutions raised a significant philosophical dilemma. Could we make these alterations without compromising the historical authenticity of the result and, with it, the mission of proving that Babbage's engines were logically and practically sound? We solved this problem by adhering to Babbage's own design practices and strictly confining ourselves to techniques or devices available to Babbage. We also planned the revisions to Babbage's design so that every mechanism we added could be easily removed.

In 1989 we built a small trial assembly at the Science Museum to verify the logic of the basic adding element and to confirm that the carry mechanism operated correctly. The assembly adds a two-digit number to another two-digit number and takes account of any carry from units to tens and from tens to hundreds. The finely finished device went a long way toward convincing sponsors and colleagues that our project involved an engineering aesthetic as well as an intriguing historical thesis. The trial piece later proved an invaluable aid for visualizing the machine's operation and for testing the first sample parts.

To build Difference Engine No. 2 and to estimate the cost of manufacturing it, we needed full-dimension drawings of its parts. Late in 1989 we contracted a specialist engineering company to produce a set of drawings using Babbage's original set as the authoritative source. Missing information—detailed dimensions, choice of materials, tolerances, methods of manufacture and a great deal of fine detail—had to be supplied.

Dimensions for the individual parts were obtained by measuring and scaling the original plans. The engineering company produced 50 new drawings that fully specified each of the engine's 4,000 parts. Surviving mechanical assemblies show that Babbage constructed his parts from bronze, cast iron and steel. Bromley and Michael Wright of the Science Museum offered advice regarding which material to use for each part. Our colleagues at the Imperial College of Science and Technology analyzed the composition of the components of Difference Engine No. 1 to guide us in selecting an appropriate modern bronze.

No attempt was made to use period machinery in the manufacture of parts. The engine's 4,000 components embody only about 1,000 different part designs, so there is a high degree of repetition. We unashamedly relied on modern manufacturing techniques to produce the many identical parts. We also welded parts that Babbage would have forged. But we scrupulously ensured that Babbage could have produced components of the same precision, though possibly by other means.

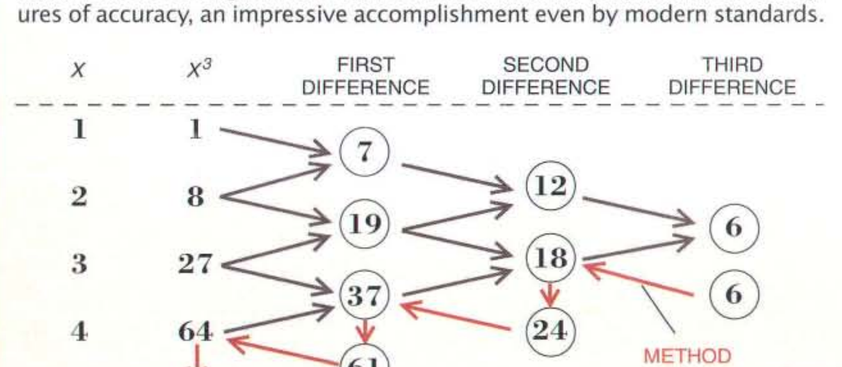
Specifying the precision with which parts should be made proved less problematic than we first feared. Bromley and Wright had measured parts from Difference Engine No. 1 and found that Clement achieved repeatability of 1.5 to 2.0 thousandths of an inch, belying the popular belief that mid-19th century mechanical engineering lacked the precision necessary for building Babbage's devices. We adopted a modern engineering standard, confident that it was within the limits of what 19th-century craftsmen could achieve. The process of producing the 50 modern mechanical drawings took about six months and was substantially complete by January 1990.

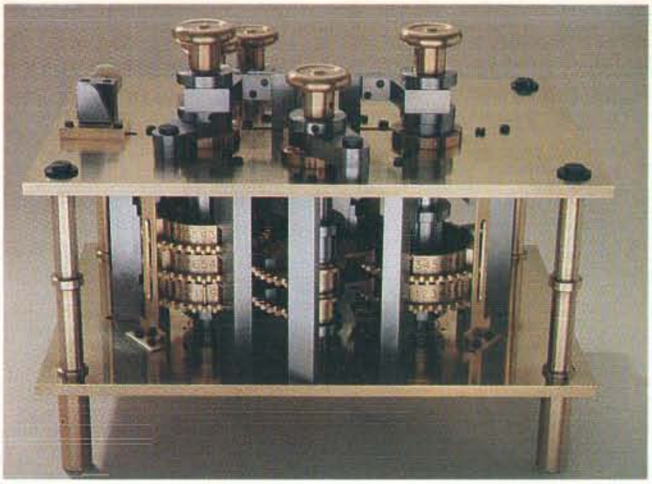
We were determined to secure a fixed-price contract for manufacture and assembly so as not to repeat Babbage's sorry tale of open-ended expense. After some hard negotiation, the Science Museum and the specialist company

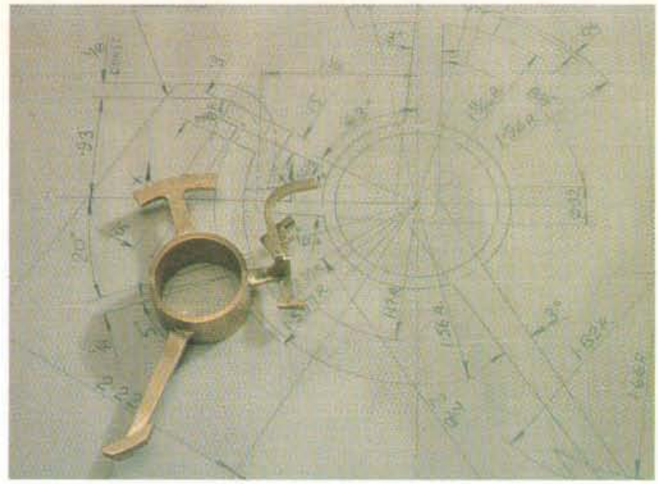
Mathematical Principles of the Difference Engines

Babbage's difference engines are so called because they use the method of finite differences to find the value of certain mathematical expressions. The method is used below to produce the table of cubes ($y = x^3$). The first difference is found by subtracting successive pairs of cubes. The same procedure is applied to pairs of first differences to derive second differences. When the process is repeated for the second differences, one finds that the third difference is constant and equal to six. This information makes it possible to generate the rest of the table of cubes by reversing the differencing procedure. For example, adding six to the second difference (18) gives the new second difference (24); adding this to the first difference (37) yields the new first difference (61). Finally, adding this to the last cubed number (64) gives the next number in the sequence, 125 or 5^3. The procedure can be repeated indefinitely to generate as many terms as desired using only repeated additions.

The method of differences can be applied to any of the mathematical functions known as polynomials, which have the generic form $y = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$. The nth difference of an nth-order polynomial will always be a constant that can form the basis for the method of differences. Polynomials are used to represent many relations in physics and engineering. They can also be used to approximate other functions, such as logarithms and trigonometric functions. In Babbage's difference engines, each column of figure wheels represents the position of one multidigit number in the table. Difference Engine No. 2 can tabulate 7th-order polynomials to 31 figures of accuracy, an impressive accomplishment even by modern standards.







ENGINEERING CHALLENGES were solved in the course of building Difference Engine No. 2. Engineers at the Science Museum constructed a part of the calculating mechanism (*left*) in 1989 to verify the design of the engine's basic adding element. They also built 210 intricate bronze levers (*right, shown atop a design drawing*) for the carriage mechanism.

agreed to a price and to a set of provisions to cushion against unforeseen technical difficulties. The Science Museum committed to underwrite the costs against pledges from a group of five sponsoring computer companies: ICL, Hewlett Packard, Rank Xerox, Siemens Nixdorf and Unisys.

Then, in June 1990, just as the final contract was about to be signed, the company involved went bankrupt after 35 years in business. Reg Crick and Barrie Holloway, the two engineers on the Babbage project, were fired on Thursday, June 7. Unless orders were placed with contractors by close of business the following day, we would incur cost penalties and have to embark on another round of financial negotiation, which would have jeopardized our goal of completing the project in time for the Babbage bicentenary. Officials at the Science Museum interviewed Crick and Holloway on the morning of June 8; by lunchtime they were museum employees. We spent the day frantically writing out part orders for subcontractors and drafting contract terms. At 5:30 P.M., I sprinted to the post office to mail the drawings and orders to the component manufacturers. We made the deadline by minutes.

Difference Engine No. 2 was built in public view in the Science Museum. Fitting and assembly commenced in November and was completed in May 1991. The engine became the centerpiece in the exhibition *Making the Difference: Charles Babbage and the Birth of the Computer,* which opened on June 27, 1991. Even then, the project kept us on tenterhooks. The three-ton Difference Engine No. 2 had not yet performed a full calculation, and it kept jamming unaccountably. We developed debugging techniques to track the source of the jams and continued to work on the machine during the exhibition. On November 29, 1991, less than a month before Babbage's 200th birthday, the machine completed its first full-scale successful calculation. It produced the first 100 values in the table of powers of seven and has functioned without error ever since. The engine ended up costing just under £300,000 ($500,000).

Our project illuminated several aspects of Babbage's skills as a designer and engineer. Historians of technology have debated whether the high standards of precision that Babbage demanded were necessary or were the product of misguided perfectionism. Some researchers have pointed out that cruder engines had been built to good effect. Georg and Edvard Scheutz, a Swedish father-and-son team who were inspired by an account of Babbage's work, built three difference engines, mostly of their own design. The first of these, completed in 1843, had a wood frame and was made using simple hand tools and a primitive lathe. Despite its comparatively rough construction, the Scheutzes' machine performed successfully before the Swedish Royal Academy.

Babbage's difference engines were larger and more sophisticated than those attempted by the Scheutzes, however. Our experiences constructing Difference Engine No. 2 underscored the importance of exacting standards. We had expected that repeat parts made using computer-controlled machines would be sufficiently identical to be interchangeable. This proved not to be the case. Fine tweaking of components to tolerances of no more than a few thousandths of an inch proved necessary, especially for the proper operation of the carry mechanism. Babbage's insistence on high precision was evidently based on sound engineering judgment.

Constructing Difference Engine No. 2 revealed subtleties and ingenuity in Babbage's design not immediately evident in the drawings. The project also gave us tremendous respect for Babbage's ability to visualize the operation of complex mechanisms without the aid of physical models. We hope to extend our explorations of Babbage's elegant designs; to do so, we are currently trying to attract sponsorship to build the printer. In the meantime, we marvel at the physical realization of plans that Babbage drew up nearly 150 years ago. Difference Engine No. 2 stands as a splendid piece of engineering sculpture, a monument to the rigorous logic of its inventor.

FURTHER READING

CHARLES BABBAGE: PIONEER OF THE COMPUTER. Anthony Hyman. Princeton University Press, 1982.

DIFFERENCE AND ANALYTICAL ENGINES. Allan G. Bromley in *Computing before Computers.* Edited by William Aspray. Iowa State University Press, 1990.

GLORY AND FAILURE: THE DIFFERENCE ENGINES OF JOHANN MÜLLER, CHARLES BABBAGE AND GEORG AND EDVARD SCHEUTZ. Michael Lindgren. MIT Press, 1990.

CHARLES BABBAGE AND HIS CALCULATING ENGINES. Doron Swade. Science Museum, London, 1991.

A MODERN DIFFERENCE ENGINE: SOFTWARE SIMULATORS FOR CHARLES BABBAGE'S DIFFERENCE ENGINE NO. 2. James Donnelly. Armstrong Publishing Company, 1992.

SELLING TO SURVIVE

by Tim Beardsley, *staff writer*

For years, the I. V. Kurchatov Institute of Atomic Energy, a sprawling complex employing 10,000 scientists and engineers on the outskirts of Moscow, was strictly off-limits to foreigners—and to the vast majority of Soviet citizens. Yet recently it took only a phone call to Semion D. Malkin, a department chief at this key military research center of the former Soviet Union, to allow a reporter to be whisked past the guards and driven through a snow-covered maze of dilapidated laboratories to his office. (Malkin's assistant cheerfully explained that he told the guards I was part of a visiting technical delegation from Czechoslovakia.)

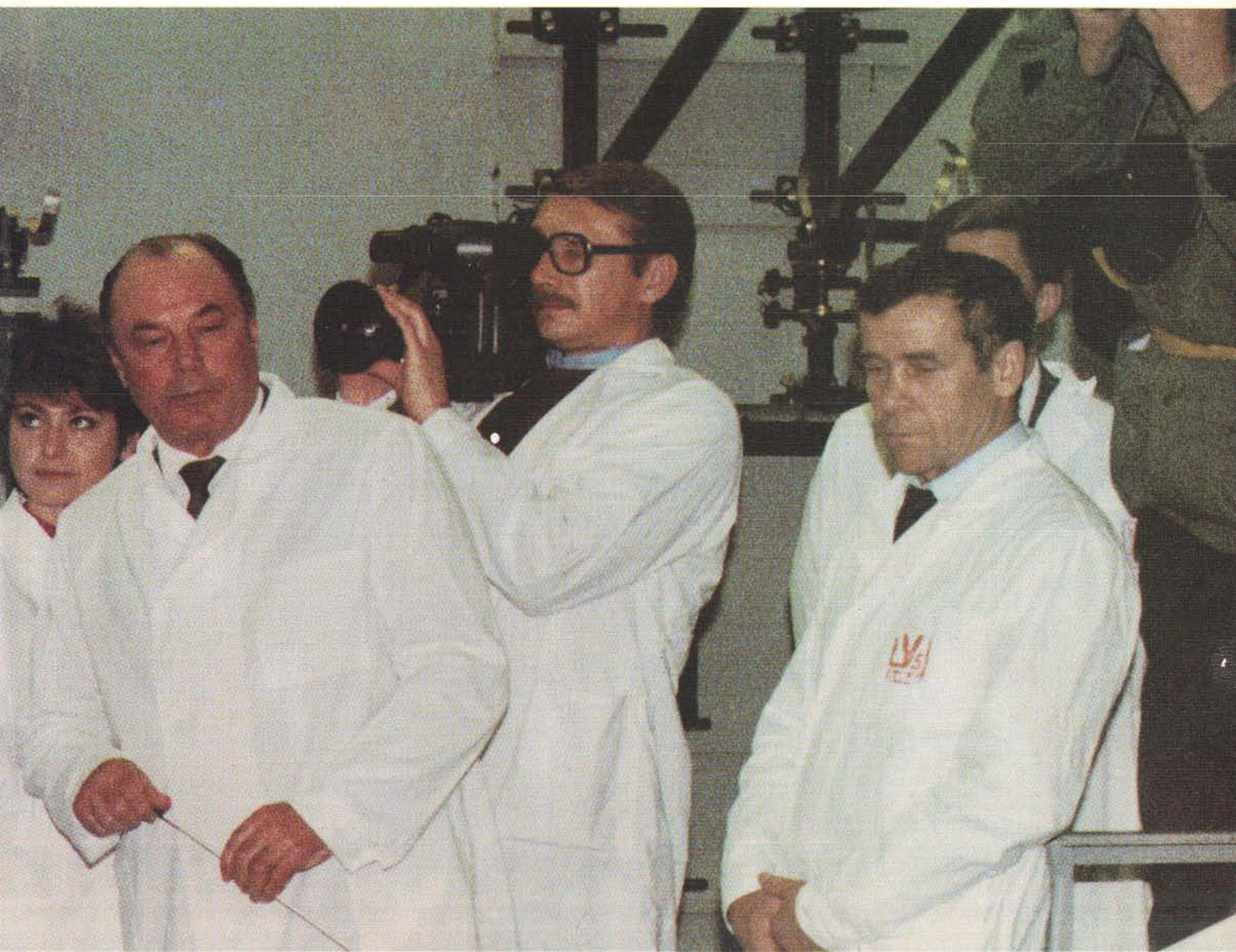
But any thought that I was the first Westerner to visit this forbidden enclave quickly vanished. Outside a second-floor office in building 22B is a nameplate in English reading "Interpatent." The words might as well be "For Sale." Interpatent, a joint venture with National Patent Development Corporation of New York, has won marketing rights to a cornucopia of previously classified technologies developed at Kurchatov and other once secret research centers.

Researchers in the Russian Federation are in desperate straits. Plummeting budgets and pitiful salaries are driving many to leave the country. Those who stay are being forced to become merchant adventurers.

Although it has the disturbing ring of a flea market at Los Alamos, these days the selling of technology by the laboratories of the former Evil Empire is neither a secret nor a surprise. Interpatent is one of hundreds of ventures created by desperate researchers in Russia, where about 95 percent of the 1.5 million scientists and engineers in the Soviet Union worked. The system of guaranteed support that established military might and technological prowess during the cold war has virtually collapsed. In what is arguably the world's largest scientific and technical work force—half as big again as that of the U.S.—researchers are mostly fending for themselves. Some are even being laid off, an almost unheard-of situation in a country that prided itself on low unemployment.

With inflation surging at 25 percent a month late last year, scientific workers were learning

RUSSIAN PRESIDENT Boris N. Yeltsin speaks to scientists and engineers at Arzamas 16, one of the main nuclear weapons laboratories in the former Soviet Union. With military support evaporating, workers at the complex, 250 miles east of Moscow, are seeking help to change over to nonmilitary production.



that the high status they enjoyed in Soviet society no longer provides a livable income or decent working conditions. Some, like those at Kurchatov, are founding companies and forging deals with the West. Others are simply fleeing, taking temporary fellowships in Europe, Israel, Japan and the U.S. Many are likely not to return.

"The new government has no conception of the management of science," charges Aleksandr V. Yevseenko, an economist at the Institute of Economics in Akademgorodok, the "science city" near Novosibirsk in Siberia. Igor A. Nikolaev, head of science and technology policy at the Ministry of Science, Higher Education and Technical Policy in Moscow, predicts that unless the government takes urgent steps to help researchers, "in probably one year the result will be a catastrophe, the disintegration of science here."

Unfortunately for science, long-term activities such as research rank low on the list of priorities for the beleaguered government of Boris N. Yeltsin as it battles against opponents of free-market reforms. Nikolaev estimates that government funding for science has fallen from almost 6 percent of national income three years ago to 1.9 percent now. Worst off are the scientific institutes that did a high proportion of defense work; according to one estimate, military research was cut by 80 percent between 1991 and 1992.

Old privilege and new reality contrast poignantly at the Institute of Semiconductor Physics in Akademgorodok. Huge letters set across the top of the institute proclaim, "The goal of science is to serve the people." But Edvard V. Skubnevsky, the deputy director, says support from industry and the government-funded Russian Academy of Sciences has fallen by more than half in the past two years. "We still have not had our payments for September," he said in November. "There is a permanent delay," he added wryly.

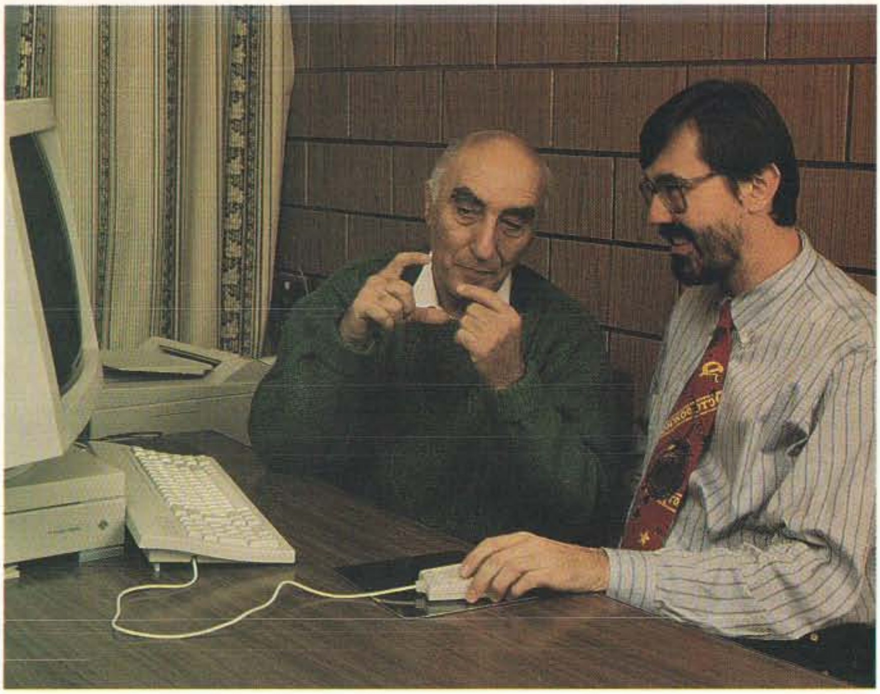
The dismal truth for a mid-level Russian researcher at the end of last year was a salary of perhaps 6,000 rubles a month, less than $15 U.S. Although living costs are low, that salary is about half what a bus driver would take home and barely enough to live on. "You have to spend most of your salary on food," says Zhores I. Alferov, director of the Ioffe Physico-Technical Institute in St. Petersburg. Even though scientific salaries have increased 10-fold within a year, food has gone up about 50-fold, he notes. Those telling figures present Russian researchers with a stark choice. "There are only two options for us to survive as scientists: to go to the U.S. or Europe or to be funded from abroad," says Maxim Frank-Kamenetsky, a researcher at Moscow's Institute of Molecular Genetics.

For researchers who have skills or knowledge that can be turned to commercial advantage, ventures such as Interpatent are one solution. Kurchatov has an equal stake with National Patent, a company with a 30-year history of successful technology transfer from what was the Eastern bloc. Malkin hopes the sale of technology will earn inventors enough compensation to keep them in Russia. "I think we have a head start," says Martin M. Pollak, Interpatent's president and a co-founder of National Patent. "Next year we'll have success in at least one area."

First for the market is a low-temperature technique for diamond coating, invented by Vyacheslav M. Golyanov. It was developed to improve the accuracy of gyroscopes in missiles, but Interpatent now wants to sell it to make protective coverings for hard and floppy computer disks as well as for bone implants and surgical thread. Next are "supersafe" nuclear reactors developed from military models that supposedly will run for 25 years without maintenance. Then come artificial lungs that work without bottled oxygen, large crystals of molybdenum, and more.

DIAMOND COATINGS for computer disks and other products are being offered to the West by Semion D. Malkin, the Kurchatov Institute's liaison with National Patent Development Corporation of New York, and inventor Vyacheslav M. Golyanov.

SUPERCOMPUTER DESIGNER Boris A. Babaian worked for the Soviet military before he teamed up with David R. Ditzel of Sun Microsystems. Babaian's team now uses Sun machines and earns dollars but will get no royalties.



Despite the promise of ventures such as Interpatent, most have yet to generate significant revenues. In the meantime, many thousands of scientists are still relying on government support, much of it channeled through the Russian Academy of Sciences. The academy, a reincarnation of the former U.S.S.R. Academy of Sciences, allocates research funds among its component institutes. But relations between the government and the academy are tense. The academy failed to condemn the August 1991 putsch that Yeltsin successfully opposed, a much remarked omission that has not endeared it to Yeltsin's team. Frank-Kamenetsky says the academy is now "the most conservative wing of society, allied with those trying to reverse reforms."

The old guard did manage to derail at least one initiative favored by Nikolaev, a science foundation that was to award grants by competitive review rather than through the academy's old boys' network. Although Nikolaev says the government intends to be more selective in how it supports science in the future, it is meeting resistance.

Exodus of Excellence

By failing to increase the academy's funding in line with inflation, the government has started to diminish the academy's influence, according to Frank-Kamenetsky. But the toll in disaffection among researchers is heavy. Yevseenko maintains that the government "is trying to oppress scientists by providing tiny salaries." More than 10 percent of the researchers at the respected Lebedev Institute of Physics in Moscow have lost their jobs.

Staff there and at other academy institutes are rebelling. The new Trade Union of Personnel of the Russian Academy of Sciences has rallied in at least two cities, protesting plans to reduce staff at some institutes by, it says, 30 percent. According to Alexei Zakharov, a union organizer in Moscow, "the government doesn't give the academy money because the academy hasn't done anything for scientists." The academy's top brass "live more like millionaires than like scientists," Zakharov says.

Meanwhile bench scientists who manage to keep their jobs often work with outdated equipment and are plagued by chronic shortages. Alferov estimates that materials have increased in cost 20- to 30-fold in the past year. Foreign scientific journals would have disappeared but for the philanthropy of Western scientific societies.

The steady flow of scientists overseas is keenly felt. Academician Nikolai N. Nikolsky, director of the Institute of Cytology in St. Petersburg, says one fifth of the 250 scientists on his staff—including many of the most talented—are now abroad. The typical stay is two or three years, although Nikolsky and other directors fear that the best will not return unless conditions improve. One survey last year counted 3,000 senior Russian scientists working in the U.S. "The U.S. is becoming home to the most luminous collection of expatriate scholars since World War II," observes Irving A. Lerch, director of international affairs at the American Physical Society.

The brain drain is most evident in institutes that emphasize theoretical work, because theoreticians can take their work anywhere. At the Budker Institute of Nuclear Physics in Akademgorodok, "we now have difficulties providing lectures in theoretical physics, as most of our theoreticians are in the U.S.," says Veniamin A. Sidorov, the deputy director.

But even institutes that are well equipped by Western standards are losing researchers. Forty-nine of the 300 scientists at the Engelhardt Institute of Molecular Biology in Moscow have gone abroad for more than a year, according to its director, academician Andrei D. Mirzabekov. "If we cannot provide facilities and salaries here, most of them will go," he predicts, adding, "I fear there will be nothing for them to come back to."

For those who are still attempting to do science in Russia, a source of hard currency has become essential. Rubles are now worth so little—a dollar fetched 450 of them in early December—that they are almost useless for obtaining supplies from abroad. Many universities and institutes run by the academy and by the government now provide little more than heat and power, researchers report.

As a result, anybody with anything he or she can sell to the West is going into business. New political freedoms together with the urgent neediness have spawned a bewildering variety of commercial entities that are vying to earn hard currency for scientists. The ventures take many forms and operate at different levels, from official institute-run concerns to under-the-desk "companies."

"There's more brainpower than you'd ever imagined," says John W. Kiser III, chairman of Kiser Research, a Washington, D.C., firm that has brokered Russian technology for seven years. Although consumer technologies lag those in the West, Kiser and others identify several industrial areas where Russia leads—welding, chemical processing, high-powered electrical equipment and metallurgy, for example. Russia also has outstanding talent in mathematics and theoretical physics. "They have done a lot of very good work and have original ideas and products," Kiser says.

RESEARCH LEADER Eugeniyi M. Dianov of the General Physics Institute in Moscow says, "I will insist on increasing salaries" when his team's contract with AT&T Bell Laboratories comes up for renewal later this year.

Recently Kiser arranged a technology transfer between the Fakel Experimental Design Bureau in Kaliningrad and Space Systems/Loral in California, a subsidiary of Loral Corporation. The agreement established a new company that will develop electrical thrusters for spacecraft based on Russian-designed "Hall thrusters." The devices could reduce the fuel consumption of satellites and so extend their lifetimes. Kiser has also interested United Technologies in a "solid flame" high-temperature ceramic processing technique. The UT research center in East Hartford, Conn., is now funding a joint research program on solid-flame technology at the Institute of Structural Macrokinetics in Chernogolovka.

Some farsighted Russian research institutes turned to the West for hard currency even before the first glimmers of *perestroika*. At the Budker Institute of Nuclear Physics, selling equipment is already a well-established sideline. The institute is known for its "wiggler" magnets, which are used in synchrotrons worldwide to produce radiation from orbiting electrons. The institute also is supplying magnets and other components for the Superconducting Super Collider, the giant particle accelerator now under construction in Texas, and has developed an innovative high-powered electron accelerator that can be used for reducing sulfur dioxide levels in flue gases and other processes.

The early pioneers in commercialism now have avid followers. At St. Petersburg Technical University, Boris V. Kuteev has created a company, Applied Physics, that is selling hydrogen pellet injectors and plasma diagnostics for machines used in fusion research called tokomaks. Kuteev has secured contracts with research facilities in China and Mexico and is lending special photographic gear to a tokomak in Garching, Germany, in the hope that researchers there will buy.

One laboratory administrator is setting up a miniconglomerate. Academician Yuri A. Ossipyan, a former science adviser to Mikhail S. Gorbachev and now a vice president of the Russian Academy of Sciences, is director of the Institute of Solid State Physics in Chernogolovka. That city of 20,000 people, built around a collection of research institutes 30 miles outside Moscow, is now facing serious cuts. Ossipyan points out that unlike residents of Moscow, those in Chernogolovka cannot simply move to the next factory. But because the institutes there were built as self-sufficient applied research centers, they have their own manufacturing plants, some of which Ossipyan has turned over to four new companies. They will produce and sell, he says, amorphous metals, inexpensive solar cells, large crystals and single crystalline metal foils. And Ossipyan is planning at least three more companies.

Even the staid Russian Academy of Sciences has joined the fray by setting up a U.S. company to broker its technology. The firm, Russian-American Science, has already established agreements with Science Applications International Corporation (SAIC) in McLean, Va., in such areas as acoustic sensors, specialized high-powered generators and environmental remediation. William L. Chadsey, a senior vice president in charge of SAIC's defense technology group, is seeking private and U.S. government funds to commercialize the technologies. "All we need is the money," Chadsey says.

Researchers for Hire

But not everyone has the marketable technology needed to form companies or joint ventures. As a result, many groups of investigators simply strike research-for-cash deals. Because of the slide of the ruble, foreign corporations are able to hire top talent for contract research at bargain-basement salaries. At the General Physics Institute in Moscow, for example, 100 researchers in the fiber-optics department under Eugeniyi M. Dianov are now supported in part by AT&T Bell Laboratories.

Dianov's group piqued the interest of AT&T when it found a way to protect and strengthen glass fibers by coating them with metal. In addition, Dianov's theoreticians lead the world in analytic studies of solitons—permanent waves

that do not lose energy and might one day be used in long-distance light-wave telecommunications.

Researchers in Dianov's group are paid an additional salary that is minuscule by Western standards—about $60 a month—but more than most would receive otherwise. The AT&T money has also meant that Dianov's laboratory can refurbish with modern equipment from abroad. Kumar Patel, executive director of the materials science and engineering research division at Bell Labs, says the interests of the Russian researchers dovetailed easily with those of AT&T. The only adjustment that was necessary, he observes, is that "we were saying to go further, beyond the proof of principle, and show a specific commercial application." Bell Labs has since hired workers at other Russian laboratories.

Researchers under Boris A. Babaian of the Institute of Precision Mechanics and Computer Technology in Moscow, a supercomputer designer, are now working for Sun Microsystems in Mountain View, Calif., on computer design and parallel-processing software. Babaian is the inventor of the Elbrus-3 supercomputer, a design that he maintains can solve problems twice as fast as can the Cray YMP, one of the fastest machines in the world.

But Babaian, who developed high-speed "reduced instruction set" techniques independently of comparable work at Sun, had been unable to put his latest ideas into practice because Russian semiconductor chips are not good enough. Now his teams in Moscow, St. Petersburg and Novosibirsk work using high-performance Sun workstations. David R. Ditzel, director of Sun's advanced systems group, says the work of the Russian team should find its way into Sun products within 18 months. Sun pays the researchers "several times" the 4,000 rubles a month they would otherwise earn, Ditzel explains. (Others say scientists in the group get about $100 a month from Sun.)

These and similar, less publicized, arrangements are now supplementing the incomes of thousands of Russian workers. But the research-for-cash deals are not without critics. Russian scientists are keenly aware that they are working for far less than they could earn in the West. And some observers suggest that commercial naïveté may have led Russian scientists to sign away reasonable rights.

One contract, an agreement between a commercial division of the Kurchatov Institute and General Atomics in San Diego, has been criticized in the Russian press. Yuri V. Esipchuk and Ksenja A. Ruzumova, two scientists at Kurchatov, are experimenting with Russian plasma-heating devices called gyratrons on a tokomak there. They are sharing their preliminary results with General Atomics, which has a contract with the U.S. Department of Energy for tokomak research.

General Atomics has a subcontract with a Kurchatov team for $90,000, which covers the work of more than 100 researchers for a year. In the U.S., that amount would barely cover one researcher. Yet the Russians say they are happy with the arrangement, in part because they benefit from seeing early results produced by "our friends at General Atomics." "This is absolutely unfair," Frank-Kamenetsky protests. "How could they be so stupid?" He suggests that if Russian scientists are treated as less than full employees of a company—which to him means being paid less than half a normal salary—they should negotiate for a share of royalties.

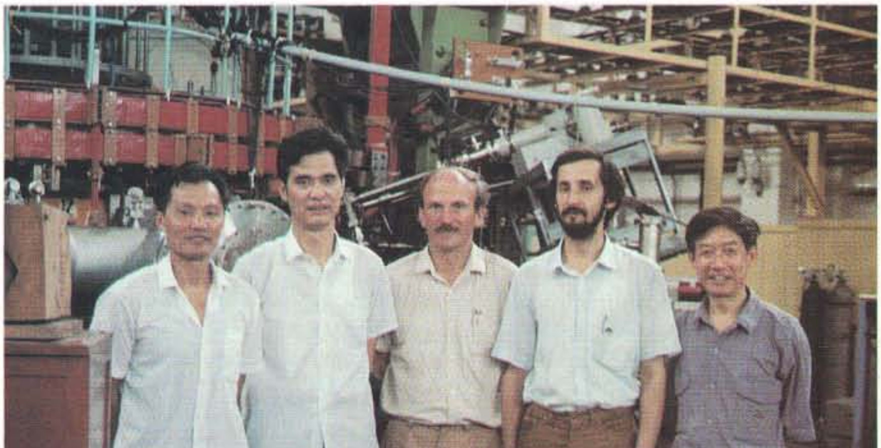
Royalty Race

Indeed, many of the contract research agreements now being struck skate over intellectual property rights, which some legal experts warn could cause problems down the road. Sun, for example, has "nothing set up now" to provide royalties for its Russian employees, according to Ditzel. He asserts that Babaian's team is treated like other Sun employees, who are rewarded for good work with salary bonuses.

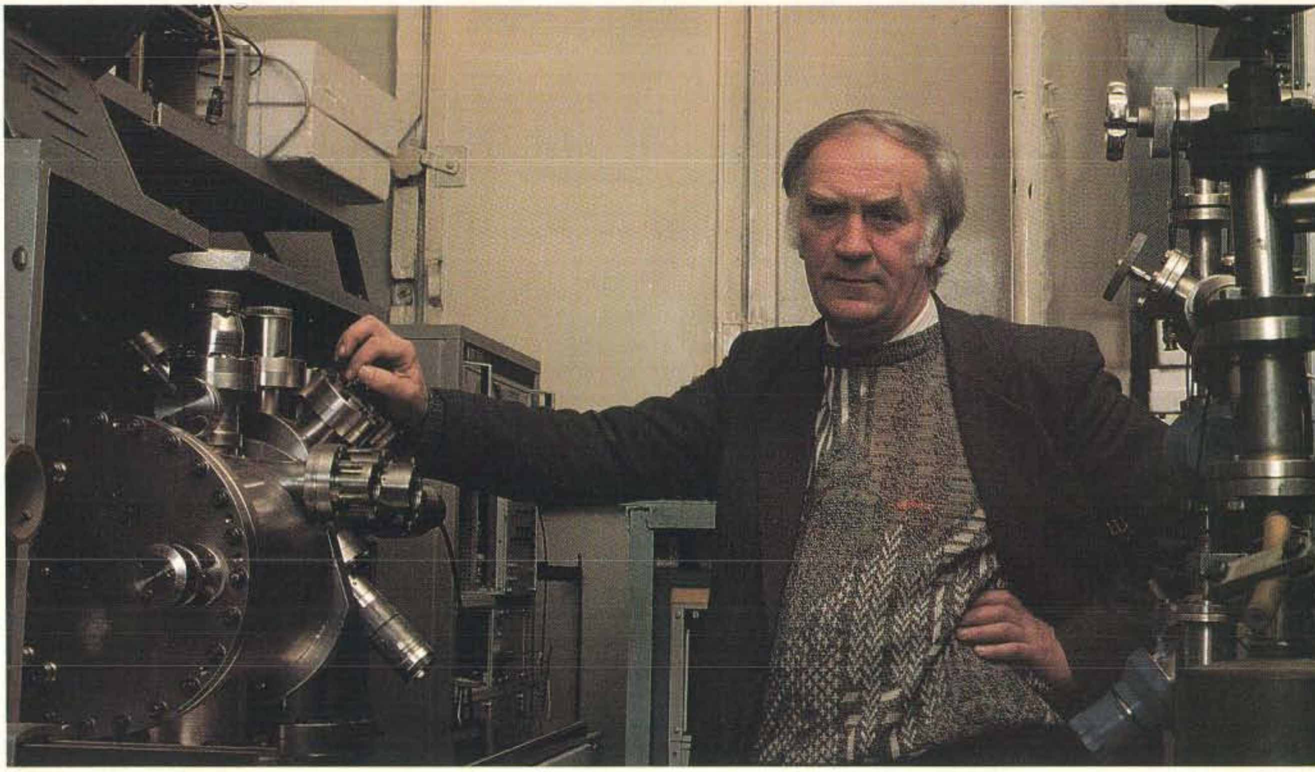
AT&T's contract with Dianov's group at the General Physics Institute is more typical. It states that any international patents will belong to AT&T, while Dianov's institute will retain ownership of Russian patents. AT&T will have the right to a royalty-free license to exploit any Russian patents. But only last September did Russia enact patent laws that brought its system into conformity with Western conventions. So U.S. lawyers are not yet sure whether Russian patents will be enforceable. The legal status of inventions registered in Russia before the new law came into effect is even more murky.

Dianov now says he is unhappy with the AT&T contract, which was his first experience negotiating a commercial research agreement. He promises that he will seek larger salaries for his group and possibly a share of worldwide royalties when the contract comes up for renegotiation later this year.

For some researchers, though, Russia's economic disarray makes being able to draw a salary and obtain equipment the main consideration. The huge Vavilov State Optical Institute in St. Petersburg, which employs more than 5,000 people, is not lacking experience with contracts. It has agreements with European companies as well as with Lawrence Livermore National Laboratory and Lockheed. Yet Gury T. Petrovsky, director of one of the institute's divisions, is happy to have secured just Russian patent rights in a recent contract that his division and the Institute for Silicate Chemistry signed with Corning. The deal covers about 100 scientists, who will receive salary supplements, and entails basic research on glasses and optical devices. "Our people can't complain about the contract," he says. "Our own government pays them less and doesn't appreciate scientists."



CHINESE PLASMA PHYSICISTS met with Boris V. Kuteev and a colleague (*third and fourth from left*) at the Southwest Institute of Physics in Leshan, Szechuan province. The Russians came away with a contract to supply equipment.



RUSSIAN PHYSICAL SOCIETY, headed by Vitaly V. Mikhailin, favors using peer review to allocate research funds. Mikhailin says the older, more powerful Russian Academy of Sciences "should be decreased by half."

Some Russian researchers are nevertheless forging agreements that they hope will serve them better. Mirzabekov of the Molecular Biology Institute proposes that if a company sponsors research at an institute, the company should be entitled only to "first refusal" of rights to exploit discoveries that may arise unexpectedly. If a company wants all rights, Mirzabekov adds, it should buy the institute itself first. "The disadvantage of Russian science is that we never learned about applications, and now we're doing so under severe pressure," he notes.

Mirzabekov's institute has a research contract with Affymax in Palo Alto, Calif., a biotechnology company that allows the Russian organization to retain international rights to inventions that are incidental to the research objective. Affymax would have rights to a license in exchange for royalties, but the institute could license to others. "It was important to us that the institute be left with enough technology rights to stimulate commercial development in Russia," says Kevin R. Kaster, patent counsel for Affymax.

The Boreskov Institute of Catalysis in Akademgorodok has also become a savvy dealmaker. Its staff has been doing research on industrial catalytic processes for overseas concerns for 25 years. It now has 20 contracts, some of them with "well known" U.S. companies, which have enabled the institute to double its income, says Valentin N. Parmon, a vice director. "We now have a permanent staff who are experts on patents and international agreements and can create any sort of contract," he boasts. The institute negotiates hard for royalties and patent rights, he says. Like a seasoned capitalist, Parmon declines to describe specific technologies, citing confidentiality agreements.

The institute is using some innovative financial ideas to maintain the flow of cash into its coffers. It encourages its employees to establish their own enterprises and form joint ventures with the institute. Recently it persuaded some employees to accept short-term contracts for higher pay. By so doing, it has gained flexibility and been able to minimize brain drain, Parmon says.

The institutes that are relying on entrepreneurial funding are not limited to those in technology. The Institute of Archaeology and Ethnography in Akademgorodok runs a hotel for visiting scientists and others near a popular archaeological site in the Altai Mountains. It also organizes international exhibitions of archaeological artifacts. Commercial activities of this kind provide no less than 80 percent of the institute's income, says Ismail N. Gemuev, a vice director.

Plans for international research centers—which would, not incidentally, also make money—abound. The Kurchatov Institute is planning a multilaboratory complex it calls Technopolis. A recently formed organization set up in opposition to the Russian Academy of Sciences, the Russian Physical Society, run by Vitaly V. Mikhailin at Moscow State University, is proposing an international synchrotron research center to aid the troubled Lebedev Institute.

Obstacles and Challenges

The problems of obtaining hard currency are compounded by the continually changing laws governing Western investments. Especially irksome are Russian taxes. Many foreign companies have been deterred from hiring in Russia by a 38 percent social security tax on salaries. In November, however, the Finance Ministry in Moscow ruled that scientific research would be exempted. That could be good news for scientists looking for Western cash—if the law is not interpreted too narrowly.

Another impediment to foreign investment is a requirement that half the value of dollar payments be converted immediately into rubles. In November, the government was considering requiring that 100 percent of such payments be converted, a move that entrepreneurs said would be crippling unless inflation is curbed.

Nor do Western laws and regulations

make it easy to invest in Russian technology. The Russian government has long complained about legal limitations that prevent Russian space hardware from being brought easily to the U.S. Another restriction makes it impossible for Russian rockets to launch U.S. satellites. "I think the U.S. could take steps to change the situation," says Nikolaev, the official in the Science Ministry.

Academician Konstantin V. Frolov, a vice president of the academy and head of the Mechanical Engineering Research Institute in Moscow, is courting McDonnell Douglas Aerospace in the hope of establishing collaborations on aerospace projects. For a meeting last November with a McDonnell Douglas representative, Frolov held a private dinner at the exquisite House of Scientists, where, in surroundings that make Versailles look shabby, the two toasted their agreement on a program covering aerodynamics, space structures and high-temperature ablative coatings.

Only two irritating details sullied the atmosphere of celebration: U.S. export-control laws mean McDonnell Douglas cannot provide technical data to support the collaboration, and U.S. government funds that might be tapped are still in bureaucratic limbo. Export controls have also been a nuisance for Ditzel of Sun. He says U.S. officials at one juncture told him he could not point out bugs in software produced by his Russian collaborators because U.S. security might be compromised. He refused to obey, and is still at liberty.

Other rules restrict the destinations of U.S. government research funds. Mikhail Kupriyanov is a researcher at Moscow State University working on single magnetic flux-quantum logic circuits. New techniques and materials have revived interest in these devices, which are based on Josephson junctions, for computation. Kupriyanov is collaborating with researchers at the State University of New York at Stony Brook but says he is not allowed to receive research funds because they originate from the U.S. Department of Defense.

The restrictions on exports to the former Soviet Union administered by the Paris-based Coordinating Committee on Multilateral Export Controls (COCOM) are also still an obstacle for some technologies, according to Patel of AT&T. COCOM rules are, however, now being relaxed as Russia starts to introduce its own export-control regime to limit the diffusion of military know-how.

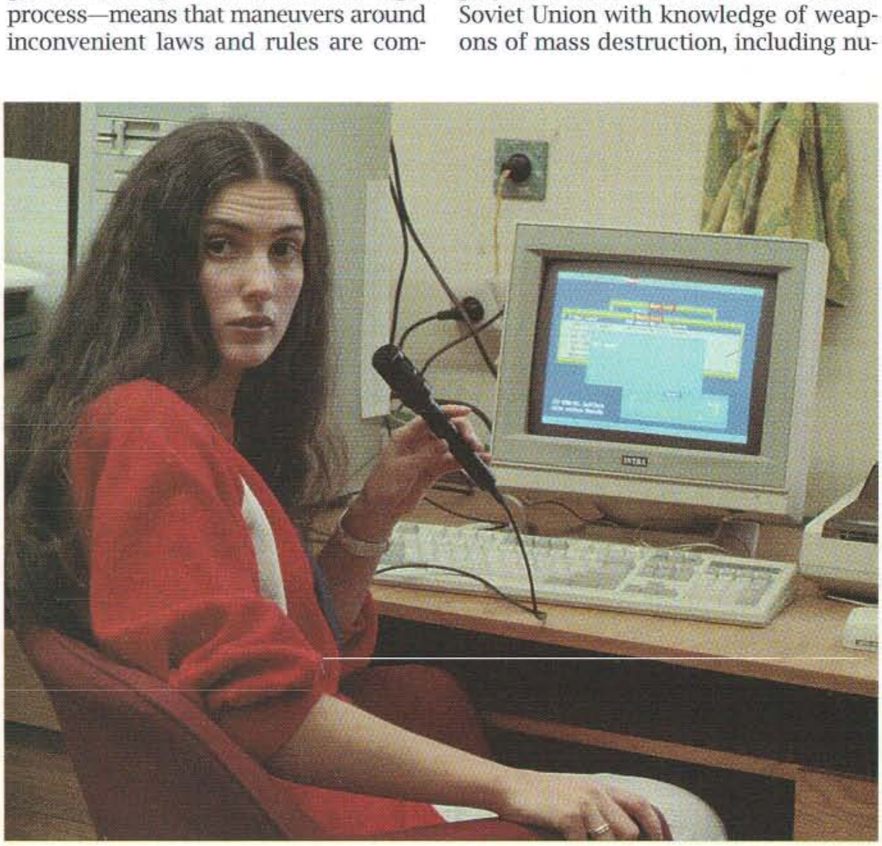
The fledgling Russian scientist-entrepreneurs who do venture into the Western marketplace complain that cold war attitudes are still reflected in unfair trade practices. Petrovsky says the Vavilov State Optical Institute has been told it often must bid at levels below world prices in order to be considered as a supplier of optical components; he reports that he declined an offer to buy biological microscopes for \$5 each. Kuteev, the physicist who runs Applied Physics, says that because his customers are not used to buying from Russians—his first visit abroad was in 1990—he has to accept 30 percent less than world prices.

Others appear to have been victims of their own inexperience. Valery I. Galloonov, director of Audio Tech in St. Petersburg, recounts how his small research-based company spent a year developing speech-recognition computer boards for a potential collaborator who was sent 400 boards. Galloonov says he has still not received payment. He and his colleague Vlasta Buzene now fear their technological lead has been lost. One of Galloonov's associates, Ivan B. Tampel, is now writing speech-recognition software for Covox in Eugene, Ore., for \$150 a month.

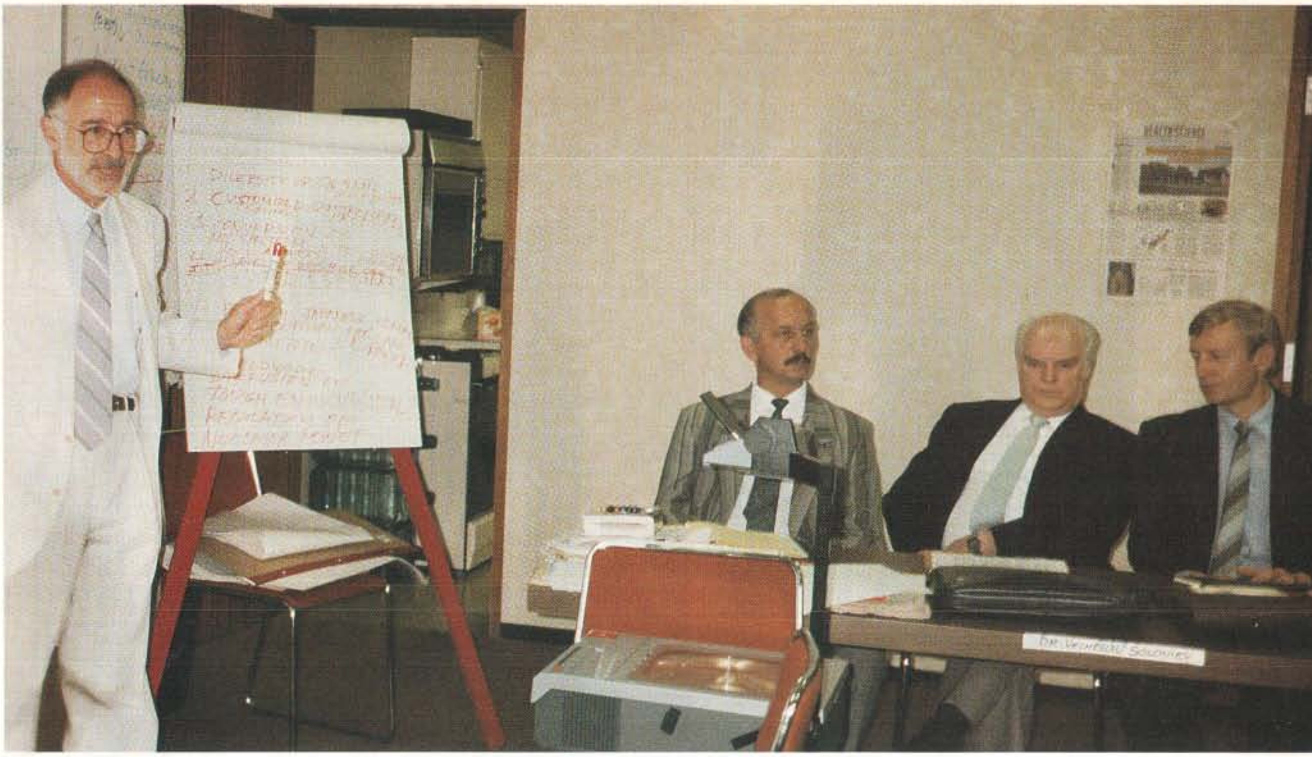
The desperate funding situation—together with cynicism about the legal process—means that maneuvers around inconvenient laws and rules are common. Indeed, much of science seems to be going underground, like the rest of the economy. One senior researcher at the Ioffe Physico-Technical Institute, for example, admits that he uses a foreign bank to hide much of his hard currency from the eyes of his boss. Otherwise, he says, he would have to turn over most of it for general laboratory support and administration. He runs a "hidden company" out of his laboratory to help support his research.

Another St. Petersburg entrepreneur, Yuri Arkhipov, is the part-time head of an analytic chemistry laboratory at the Forest Academy, a job that seems to take a backseat to his position as owner of a chemical exporting company called Science Link. Arkhipov insists his dealings are all legal, although he admits to switching money between bank accounts to minimize taxes—using a system that he calls an "account helicopter." "People have lost the feeling of community," says Anatoly Kozyrev, a senior researcher at the Central Mathematical Economic Institute in Moscow. "It's a crisis of moral institutions. Directors of institutes can no longer influence scholars, and people are not publishing their results because everybody is trying to create an enterprise."

It was fear of a moral crisis that prompted the U.S., Japan and the European Community to commit \$70 million last year to find peaceful employment for researchers in the former Soviet Union with knowledge of weapons of mass destruction, including nu-



SPEECH-RECOGNITION security system developed by Vlasta Buzene and her colleagues at Audio Tech in St. Petersburg can operate on a small personal computer. But the company is still awaiting paying customers.



WEAPONS SCIENTISTS from former top-secret defense laboratories, Arzamas 16 and Chelyabinsk 70, attended a workshop at Boston University aimed at helping them find civilian applications for their research. Western governments have committed funds to support efforts along those lines, but they have been delayed by bureaucratic red tape.

clear, biological and chemical weapons as well as ballistic missiles. Scientists from two Russian weapons laboratories, Arzamas 16 and Chelyabinsk 70, attended a workshop last summer at Boston University to learn how to prepare grant applications.

But the offices that will dispense the funds, known as the International Science and Technology Centers, were still not operational late last year, months after they had been expected to open. According to the director of the Moscow Center, Glenn E. Schweitzer, changes in Russian laws, coupled with the need to translate legal documents into all the languages of the European community, have prevented it from even considering applications.

The U.S. contribution to the International Science and Technology Centers is part of $800 million appropriated by Congress for projects aimed at making sure Soviet nuclear weapons are dismantled and the bomb makers find peaceful work. But plans have not progressed far.

The spirit of free enterprise seems to have moved faster. Police in Germany have arrested dozens of would-be smugglers who arrive from Eastern Europe carrying radioactive materials that they hope to sell. Some of the materials originated in the former Soviet Union, according to German authorities.

A few Russian scientists have criticized as demeaning Western efforts to find employment for bomb builders. But the smuggling seems to reinforce concerns about military matériel finding its way to governments with deep pockets and few scruples. Arkhipov contends that many new capitalists are getting rich by selling stolen strategic materials.

What will become of those researchers in disciplines with nothing to sell? Although it seems that any Russian scientist able to turn his or her hand to applied research is doing so for profit, such endeavors will not support fundamental research. Fortunately, efforts in the West to provide temporary support are under way, although the amounts of money are small.

The American Astronomical Society has been quietly handing out $100 bills to Russian astronomers. The National Institutes of Health have some joint grants. And the Soros Foundations, philanthropic institutions founded by Hungarian-born businessman George Soros, have teamed up with the American Physical Society, the Sloan Foundation and the National Science Foundation to provide $800,000 for science in the republics that formerly constituted the Soviet Union. The physical society distributes the funds as competitive grants to individuals. In December, Soros committed a further $100 million through the new Foundation for Science in the Former Soviet Union. It will support individuals, groups and institutions, providing emergency stipends as well as creating fellowships and endowing 100 permanent chairs.

Everyone agrees, however, that such efforts can be only a Band-Aid, temporary protection that will ultimately have to be replaced by homegrown support. In the end, whether science thrives in Russia will depend on the choices made by Russian politicians. If the government fails to pay adequate wages and makes it impossible for scientists to profit from their creativity, the brightest and best will surely leave. As the exodus from Germany before and after World War II and the British "brain drain" attest, it would not be the first time that the U.S. and other countries have benefited from a mass emigration of scientists.

The next generation would then have no teachers, and the Russian scientific tradition would be lost. Proposals for joint Western and Russian laboratories that would serve as training sites for tomorrow's scientists are now being discussed, and some models already exist. But time is short. "If people in Western countries would like Russia to be a civilized country where educated people are important," Mirzabekov says, "science here simply must be supported."

Zip Code Breakers

The Postal Service automates the reading of chicken scratch

POSTAL-CHARACTER READERS, such as the ones shown here, will be supplemented by devices that can discern handwriting and hard-to-read type on envelopes.

I just finished addressing an envelope to myself. The scratchy letters resembled more the blips of an electrocardiogram than handwritten script. The scanning and recognition machine from the State University of New York at Buffalo thought so, too. First, it was tripped up by the bouncing array of happy faces that I had scribbled all over the envelope. Then it could not deal with the pretentious, European-style, crossed seven that I had scribed on the zip code.

Mine was one of the seven out of 10 handwritten addresses that the machine's software could only partially read or else route directly toward the reject bin after staring at the words for up to a minute or so. The world's most sophisticated neural network—a grumpy human mail sorter—would have routed the envelope on its way, probably in about five seconds.

By the turn of the decade, however, the U.S. Postal Service wants more address scanning and fewer mail clerks. The goal, by then, is to distinguish one out of two handwritten addresses in a second or, preferably, much less time. The reason is simple: it costs \$40 to sort 1,000 letters by hand, an amount that drops to \$4 if done by machine. Will my handwriting get better, or will S.U.N.Y.'s algorithms read more accurately? Leading academic and corporate representatives who attended a Postal Service conference from November 30 to December 2 in Washington, D.C., are betting on the latter.

The sheer volume of mail defines the magnitude of the problem facing the Post Office. The service now processes 550 million letters every day, about 85 percent of which have a machine-printed address; the rest are handwritten. Half of the mail with machine-printed addresses gets correctly sorted by optical character-reading (OCR) machines that operate at speeds up to 13 envelopes a second. The goal is to increase the accuracy rate to 90 percent for envelopes addressed with machine print and to 50 percent for the handwritten mail. "The Postal Service has the most challenging OCR problem in the world today," says Charles E. Stenard, supervisor of Advanced Information Systems at AT&T Bell Laboratories. "Sponsorship of these development programs is undoubtedly driving the technology."

Last year the Postal Service entered into nine contracts whose value ranged from hundreds of thousands to millions of dollars with AT&T, IBM, Westinghouse, TRW, Hughes Aircraft and other technology firms. The goal is to develop a machine that can process both handwritten and type-printed discards from high-speed optical scanners. The intention is to allow these rejects to be stamped with a bar code for the street address, letting them be sorted without human intervention.

The production contract, estimated by one supplier to total about \$150 million, will be less than a third of the cost of a now scrapped effort to develop a new generation of high-speed OCRs that would have incorporated a more advanced version of these recognition technologies. The machine currently being developed will perform the same tasks but at slower speeds.

Postmaster General Marvin Runyon dropped research funding for the more ambitious effort (the original plan was to fund both projects) to help stave off a projected \$2-billion deficit for the current federal fiscal year. Postal officials plan to proceed with a replacement for the electronics in the current generation of OCRs, although much of the development bill may be footed by the companies that bid for the contract.

Despite the cutbacks, the Postal Service will probably continue to fund a 90-member research center at S.U.N.Y./Buffalo (and some other institutes) devoted to pattern-recognition problems, research that the companies developing the products will eventually use. The challenge of discerning an address on an envelope differs from that of enabling a computer to recognize handwriting executed with a stylus, a technology that is now commercially available. The pen device need only be trained to recognize one individual's handwriting. Any system that the Postal Service adopts must be able to work from the general case to the particular—and it lacks critical information about pressure, motion and time that is captured while a person is actually writing.

Breaking down cursive script into individual letters is one of the most dif-

ficult tasks confronted by the computer. To a computer, a handwritten, looping "l" looks suspiciously like an "e" turned on its back. For that reason, the research staff in Buffalo has gone beyond simply trying to analyze the individual characters. It attempts to capture the shape and length of the whole word, using the incidence of ascending and descending letters as one clue. "The ultimate goal of this research is to do this task as well as a human does," says Sargur N. Srihari, professor of computer science at S.U.N.Y./Buffalo and director of the Postal Service's Center of Excellence for Document Analysis and Recognition.

Another approach—looking at individual characters—is often supplemented with a technique borrowed from speech recognition, a Markov chain in which a probability is assigned to a letter based on the one that is surmised to have preceded it. "In no case can we be very confident of a single piece of information, so we have to use a number of techniques," says Venu Govindaraju, a research scientist at S.U.N.Y./Buffalo.

The standardized address format can be used to assist in reading, one advantage that pen computers do not have. Correlating a zip code with a street address number, for example, can usually narrow the search for the recipient to about 10 possibilities within the Postal Service's multigigabyte, nationwide data base. "If you know the zip and the street number, you can substantially limit the number of candidates you are faced with to just a few streets," says Carl G. O'Connor, an engineer with a Postal Service development group.

For now, the pace of recognition of most prototypes is glacial. It can take up to several seconds to tell a zip code and a minute or thereabouts to make out a full address—nowhere near the one to 13 pieces of mail per second that is the goal of the program. Meeting those objectives will probably require the use of specialized computers and algorithms that process data in parallel or that mimic the pattern-recognition abilities of the human brain.

Enter Synaptics, a Silicon Valley start-up begun by two pioneers in semiconductor technology. The company has developed neural-network chips that specialize in recognizing machine-printed addresses. It has also done research on recognizing handwritten zip codes at speeds of thousands of characters a second. (Other Postal Service contractors are pursuing similar research.)

Instead of relying on step-by-step algorithms, a neural network may help decide where things begin and end within the address block, a task that can easily stump a machine. Ralph Wolf, a Synaptics researcher, describes training the network with 800 examples of Post Office-supplied type-printed addresses with different sizes, fonts and locations on the envelope. After much exposure, the neural network recognized the location of the corners of an address on the envelope by generalizing a pattern from the hundreds of samples. But it still had difficulty when it came upon a return envelope for a religious organization with a crucifix that extended into the address area.

Deciphering crossed sevens may also take time that the Postal Service does not have. Whether the agency will be around 20 years from now is a question that even its administrators have already begun to ponder. Robert W. Lucky, vice president for applied research at Bell Communications Research, told a roomful of postal officials from all over the world at the recent conference how he had monitored his own mail for four days. In this decidedly unscientific survey, he received 37 items: nine catalogues, 14 pieces of junk mail, eight bills, three announcements, two magazines and only one personal letter, a breakdown that roughly corresponds to the numbers the agency registers for the average household.

With a mind-set for electronic information, Lucky calculated how this stream of mail would translate into bits and bytes. He noted that more than half of the bits would have been taken up by advertising catalogues. The message, Lucky says, can be summed up in a word—advertising. "If this wasn't seen as a cost-effective way to get to a house, the rug would be gone from underneath you," he told the postal officials.

Although paper is still the preferred medium for many, Lucky pointed out that the data contained in his daily mail bundle—about one gigabit a day when translated into information-age terminology—could be transmitted by fiber-optic networks that are capable of carrying up to 30 trillion bits a day to the home. This may mean a slow death for the catalogues that seem to be one of the Postal Service's mainstays. "If they could deliver the right page out of the I. Magnin catalogue to the right person, they wouldn't send the whole catalogue," he says.

Even if people will one day sit before their televisions to read a catalogue or the daily paper, the work funded by the Postal Service may be a legacy to future generations of electronic-mail carriers. There will always be some need to convert paper to electrons and vice versa. How do you say, "You may have just won $1 million" in ASCII? —*Gary Stix*

Surreal Science

Virtual reality finds a place in the classroom

High school physics might have been more fun—and more instructive—if the ability to bounce balls on Jupiter were a laboratory option. Well, now it is, at least in principle. Using his experience in developing virtual-reality environments to train National Aeronautics and Space Administration astronauts, R. Bowen Loftin, a physicist at the University of Houston-Downtown, has developed a virtual-physics laboratory in which this and other improbable exercises are as easy as making a slam-dunk shot on a wastepaper basket.

Students enter virtual reality by donning headgear and a control glove hooked up to computer. A panoramic view of the virtual laboratory is projected through small screens inside the headset. There is a table, a pendulum, some balls as well as a few odd devices that govern the actions in the laboratory. Users see a computer-generated image of their hand, which duplicates the motion of the real thing. Certain gestures of the control glove are reserved for special actions. Pointing the index finger, for example, will send the virtual physicist flying across the room.

In this artificial environment, students can conduct simple experiments in mechanics that are impossible to perform in a real world. Parameters such as friction and drag can be controlled; gravity can even be turned into a negative force. Computer-generated trajectory lines and freeze-time capability enable users to measure the effects of the changes in the variables on swinging pendulums and bouncing balls. Balls, for example, can bounce so that they increase in energy with every skip.

Amusement, however, is not Loftin's primary goal. He and other developers of educational technology are trying to create an interactive environment that challenges students' often mistaken ideas about the physical world. "Kids build misconceptions about the world that are difficult to dislodge," points out Christopher Dede, who directs the center for interactive educational technology at George Mason University.

Loftin notes that many people think constant motion always requires force. This perception—a violation of Newton's first law—comes about because friction almost always plays a role in the motion of objects. "Newtonian mechanics is not easily detected in the world around you," he says.

SCIENTIFIC AMERICAN CORRESPONDENCE

Reprints are available; to order, write Reprint Department, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111 or fax inquiries to (212) 355-0408.

Back issues: $6 each ($7 outside U.S.) prepaid. Most numbers available. Credit card (Mastercard/Visa) orders for two or more issues accepted. To order, fax (212) 355-0408.

Index of articles since 1948 available in electronic format. Write SciDex™, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111, or fax (212) 355-0408.

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN (for the fee of $3.00 per copy of each article plus $0.50 per page). Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 27 Congress Street, Salem, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/93. $3.00 + 0.50.

Editorial correspondence should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111. Manuscripts are submitted at the authors' risk and will not be returned unless accompanied by postage.

Advertising correspondence should be addressed to Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111, or fax (212) 754-1138.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 3187, Harlan, IA. 51537. The date of the last issue of your subscription appears on each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (mailing label, if possible) and your new address.

The immersive quality of virtual reality, Dede says, makes it ideal for such subjects as physics. "We can get a direct intuitive qualitative sense of physical laws. Such simulations are not possible with a mouse running around on a flat screen." Furthermore, the virtual-physics laboratory is not so dangerous as some real physics experiments, as anyone who has gotten a shock from a spark generator or has been bopped by a pendulum knows.

Although most virtual-reality systems retain a certain level of unbelievability, they have proved to be especially useful in situations that would otherwise be difficult to simulate. The near weightlessness of space is one such obvious example. The more conventional techniques all have shortcomings.

In underwater tanks, the viscosity of water affects the training. Freely falling aircraft such as the KC-135 can produce true weightlessness for only about 20 or 30 seconds. Air-bearing floors that levitate astronauts on a cushion of air do not provide the full three-dimensional motion of floating in zero gravity. Virtual reality, on the other hand, can accurately simulate all the conditions of space, so NASA is now using that technology for such tasks as training astronauts to grapple wayward satellites and use robotic arms.

But it may take a few years before such computer simulations become commonplace in the classroom. The virtual-physics laboratory is still rather rudimentary. Even by the standards of video games and television cartoons, the images have a stilted quality. Also, the computational speed limits the system's repertoire to a few simple mechanics experiments. But by March, Loftin expects to have a version that will enable users to explore such concepts as angular momentum and energy conservation.

The future could bring virtual laboratories for other domains of science, such as chemistry, biology and medicine. Indeed, Loftin says that medical schools have shown interest in virtual reality for anatomy courses. "It could be like *Fantastic Voyage,*" Loftin observes, referring to the movie based on Isaac Asimov's story of scientists shrunk and injected into a human body.

Prices, too, will have to come down before virtual laboratories become widespread. Loftin's headgear and control glove cost several hundred thousand dollars. Researchers are confident, however, that within 10 years, virtual-reality circuit boards should become available for desktop computers. Then, coupled to modest peripheral devices, the fidelity would be sufficient to do some really surreal science. *—Philip Yam*

Learning Companies

Educating corporations about how people learn

At Steelcase, Inc., in Grand Rapids, Mich., researchers are redrawing office blueprints to feature large shared spaces, or "commons," dotted with tiny spots of privacy called personal harbors. Their goal is to construct an environment in which people learn to accomplish tasks more rapidly and so work more efficiently. At American Airlines in Dallas–Fort Worth, Tex., researchers hope to find ways that allow diverse airport crews—from the ticket agents at check-in to the flight attendants on board the aircraft—to provide better service by learning to work together seamlessly.

From management wizards such as Tom Peters to pop business magazines such as *Inc.*, there is a blizzard of enthusiasm for spurring learning at work. "We're trying to unravel why, in spite of this enormous investment in computers, we have had no productivity gains to show for it," says William L. Miller, director of research and development at Steelcase. Many believe an answer may lie in building environments that support learning. Firms may then boost workers' abilities to solve problems on the fly and so innovate. But spirits sag as companies realize they know very little about what strengthens—or cripples—learning.

Consultants sell ready-made solutions for companies. Panels of experts try to fix schools. But the small, nonprofit Institute for Research on Learning (IRL) in Palo Alto, Calif., is one of the few efforts looking for the common threads in how people learn at school, work or anywhere else. "I've been studying organizational learning for five or six years now, and it seems that IRL has done the most basic, innovative and profound research in this area," declares Patricia Seybold, president of the Patricia Seybold Group, an established publishing and consulting firm in Boston.

IRL is itself an experiment. The institute, which has roots in Xerox's Palo Alto Research Center, was launched as an independent research collective in 1987 by former Xerox chairman David T. Kearns. Its montage of linguists, anthropologists, computer scientists and professional teachers began with the hunch that people learn less through formal instruction and more through social interactions. "We can't assume that one discipline [professional teaching] has a lock on learning," says Susan U. Stucky, associate director of IRL.

Lightning Lure

In February 1992 a lightning strike triggered an automatic shutdown of the 460,000-kilowatt Shimane Number One Reactor of the Central Japan Power Company. To avert another such occurrence at Shimane or any of the other 43 nuclear power plants operating in Japan, corporate laboratory researchers as well as workers at universities in Osaka, Nagoya and Kyushu have been exploring an unusual preventative.

Following the lead of experiments conducted in the U.S. and the Soviet Union 20 years ago, the Japanese investigators are attempting to develop laser-beam conduits that can guide lightning bolts safely to the ground. A laser beam would do so, the researchers suggest, by creating an ionized pathway along which a bolt would flow, like water in a pipe.

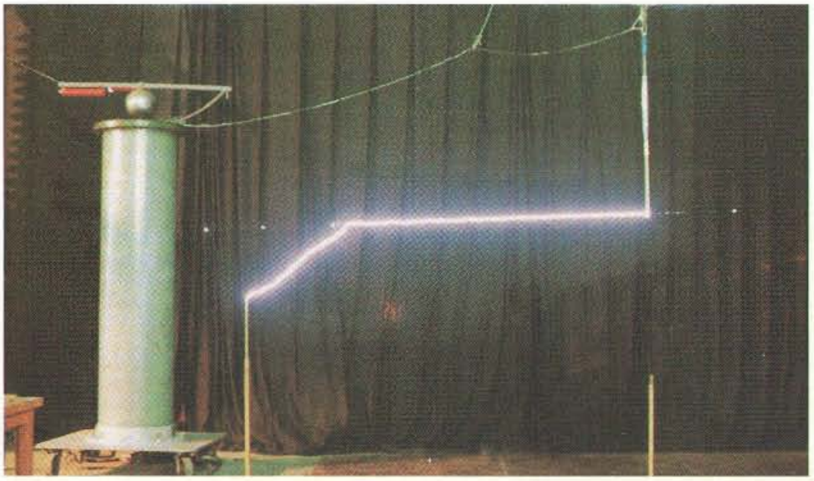
Last year a team from Osaka University, Kansai Electric Power Company and the Institute for Laser Technology successfully shepherded an artificial lightning bolt over a distance of 8.5 meters, setting a world record. At Japan Electric Power Industry's Central Research Institute near Tokyo, another group is working to beat the Kansai mark. In one of the institute's huge experiment buildings, laser-guided lightning bolts routinely zoom three meters horizontally across a copper-clad lab floor.

The bolts are produced when pulses of infrared radiation from a 300-megawatt carbon dioxide laser are directed across the gap between a hanging wire electrode and a grounded post. The laser energy heats dust and other particles in the air, vaporizing them into beads of plasma that line up like a string of minute conductors between the electrodes. Although the plasma exists for only a fraction of a second, it is time enough for a 600,000-volt streak of lightning to zap to the ground. "The lightning makes a very big sound," says Takatoshi Shindo, head of the institute's research team.

Shindo hopes to be ready later this year to move the tests to the institute's outdoor site high on a mountain plateau in Tochigi Prefecture, north of Tokyo. The facility features a 12-million-volt impulse generator, the world's most powerful. In a test conducted with the device last July, Shindo's group guided a lightning bolt over a course seven meters in length.

Now he plans to extend the distance to 20 meters or more. To this end, Shindo has been testing new cone-shaped lenses that can focus the laser light over greater distances than can concave lenses. These, along with multibeamed lasers, may be refined to extend plasmas for tens of meters. Shindo's team has also come up with a clever way to use the relatively short plasma beams produced by existing technology. The workers have designed an array of beam splitters and lenses that would produce a segmented curve of plasma reaching from the clouds to a grounded tower.

The system will first be used in winter, because the bases of storm clouds at that time of year are much closer to the ground than in summer. Even so, Shindo believes his high-tech version of Ben Franklin's kite and key will eventually lure lightning away from dangerous targets. —*Tom Koppel, Tokyo*



ARTIFICIAL LIGHTNING is guided by a laser to a grounded post at Japan Electric Power Industry's Central Research Institute near Tokyo.

Over time, the researchers (who now number more than two dozen and operate on a \$3-million budget) realized that they, too, needed to learn through working with others. So like the M. C. Escher drawing of two hands simultaneously sketching each other, IRL soon found itself both studying and experimenting with the factors that shape how people acquire knowledge.

One early institute partner was Philips Corporation of the Netherlands. "The cooperation with IRL worked extremely well for us," asserts Jaap de Hoog, a director of Philips Consumer Electronics in Eindhoven. For two years, Philips and IRL collaborated on the design of a communications device that would pack a telephone, videophone, fax, electronic mail and so on into one box likely to be used by small businesses. Philips managers knew that unless the technology seemed more accessible than a videocassette recorder, no one would buy it.

The institute researchers advanced two premises. First, that people absorb information more readily when engaged in their daily tasks than they do through special classes or by reading manuals. And second, that people grasp ideas faster when they take part in uncovering them. A device could be quite sophisticated, the scientists suggested, provided that it was designed so that people gradually learned how to use it as they did their routine work.

But there are no general principles that dictate how to design such a product. So Philips and the researchers embarked on a regime of making prototypes, putting them into real offices and recording through extensive videotaping and interviews how people grappled with the technology over time. "We spent tens of hours doing detailed analyses of those videos," says Robert de Vogel, a Philips project manager.

The methodology behind their painstaking scrutiny of actions and gestures is vintage IRL: a blend of ethnography, linguistics and intense discussion. "There's no fancy technology involved," points out Brigitte Jordan, who started the group's interaction analysis lab. But the move-by-move assessments help designers to understand how the technology appears from the users' perspectives, Jordan explains.

For financial reasons, Philips eventually decided not to build the product, de Hoog says. But the IRL-inspired approach of evaluating emerging designs continues to shape new products such as medical diagnostic displays and audiovisual equipment. Recently Philips researchers were testing a prototype of a high-end digital audio system for scanning and taping radio broadcasts. On the videotape, the person using the equipment seemed to have no trouble with it. But a careful analysis of the taped record revealed he had misunderstood how the device worked and had succeeded only through chance.

The learning institute's work is also helping Steelcase see the office in a different light. In the past, Steelcase designed offices and furniture with an eye toward enabling any one person to do his or her job. But IRL's research indicates that people learn more swiftly—and so are more effective—when they belong to "overlapping communities of practice," or groups with differing specialties, Miller says.

As a result, Steelcase's current experiments in office layouts emphasize common space rather than insular cubicles. Large, mobile easels and white boards for jotting down ideas are plentiful. In contrast, Miller notes, "a sheet of paper is optimal for one person." Personal harbors, which are small, mobile enclaves, offer spots where people can concentrate without distraction.

Early this year Steelcase plans to try out its designs in about 20 customer sites, from insurance companies to computer manufacturers. Through analytical observation Miller hopes to understand how the newly configured spaces change work habits and "to help clients learn how to learn."

Institute researchers argue that groups play an equally crucial role in schools. IRL sociolinguist Penelope Eckert has done studies that show how both high school overachievers and "burnouts," or disenfranchised teenagers, are continually learning, albeit not through classes. Instead, through their social networks, both groups of adolescents are soaking up lessons about how to interact with people, what constitutes authority and what knowledge is useful. Such lessons, she suggests, will eventually shape their success in the workplace, as much as will their knowledge of, say, mathematical fundamentals.

Other researchers are trying to catalyze the formation of communities in classrooms to help students take in the formal curriculum. In partnership with teachers in five San Francisco Bay Area schools and several scientists from Sandia National Laboratories, the institute workers are constructing a novel mathematics program.

Beginning this past September, for instance, middle school students in the experiment took on the challenge of designing an energy-efficient scientific base in Antarctica. In trying to find the best shape for their buildings, the students learned about geometry, such as how perimeters are related to volumes, according to Shelley V. Goldman, one of the IRL investigators. Moreover, "the whole package relies on kids and teachers having conversations," she says, and "co-inventing" solutions. Goldman's team also aims to co-invent the programs themselves with the teachers. By encouraging teachers to draw on local scientists to design vivid projects, "we're trying to create a model of a 'community' around the teaching materials," Goldman explains.

INSTITUTE FOR RESEARCH ON LEARNING

VIDEO of people learning to use new technology helped investigators from Philips to improve the design of an experimental multimedia communications tool.

Such evolving communities are close cousins to those that naturally arise among people who work together, the researchers observe. For instance, Mark C. Maletz, managing director of business process design at American Airlines, argues that there are many communities of practice surrounding the departure gate at an airport. Gate agents, the cockpit crew and flight attendants, and the battery of people who load the plane with food and fuel all must work together to get the flight off on time. But when a departure is delayed, Maletz says, "people ask, 'Who's the group that screwed up?' It's a win-lose situation."

A "reasonable percentage" of airline problems could be solved, Maletz suggests, if the groups are aware of how they interact and jointly try to improvise solutions. Imagine, he says, that a particular type of engine suffers from a recurring mechanical glitch during turnaround. If the crew of an incoming airplane alerts the ground crew of the likely problem, mechanics will have more time to gather the necessary tools. But spotting that pattern means workers must look at their work and the role of others differently.

"IRL has an ability to see things we can't, a set of 'lenses' that bring such aspects into focus," Maletz says. And his hope is that, through practice, American Airlines will also learn to see better ways of working. *—Elizabeth Corcoran*

Shell Shocked

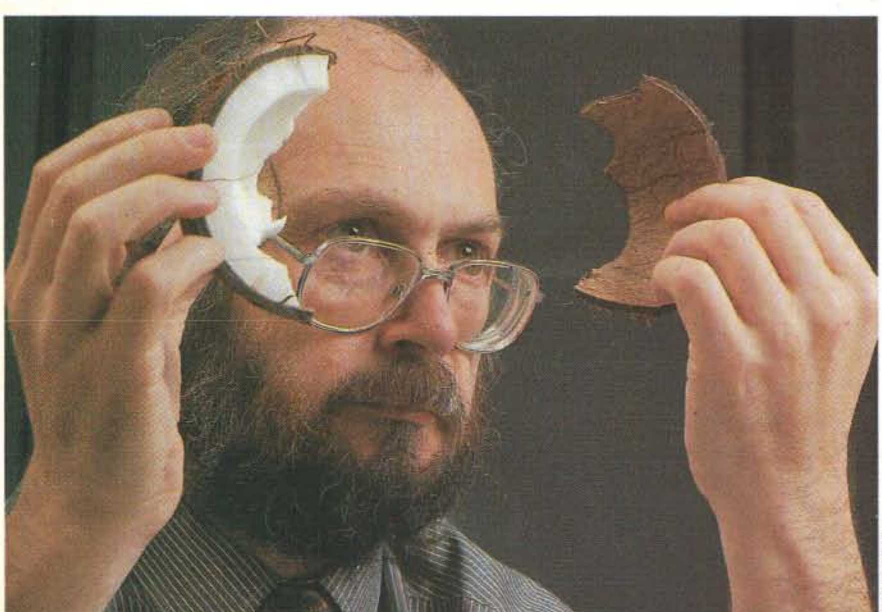
Nuts may inspire materials designers

Should the designers of aircraft wings and crash helmets take a lesson from the walnut? Julian Vincent, co-director of the Centre for Biomimetics at the University of Reading in England, thinks so. The outer layers, cells filled with a hardened, resinous substance, efficiently resist compressive stresses, while the fibrous inner layers appear to hinder tensile fracture. Yet a walnut shell is relatively thin, relying instead on a robust design. "Nature thinks that material is difficult to come by, but finding a good structure is cheap," Vincent observes.

Vincent's straight-faced presentation was tucked away among the thousands of papers delivered at the Materials Research Society's fall conference in Boston. Despite the laconic title of "Nuts" and the fact that he also recounted how he slammed a weighted pendulum into a coconut shell to test how it might serve as a helmet for skateboarders, Vincent is hardly a crackpot.

Looking to nature for design ideas—a strategy dubbed biomimetics—requires microscopic examination of everything from silk to rubberlike proteins that come from cockroaches. Vincent has distinguished himself in this subdiscipline by concentrating on how things fracture and break. He and his co-workers apply the principles of fracture mechanics to apples, potatoes, wood and animal horns, to name just a few kinds of materials.

The center is currently under contract to the British Flour Milling and Baking Research Association for a study on wheat that may help improve the efficiency of milling that grain. For the past year, the 10-member staff has also been putting potatoes under x-ray and light microscopes for Unilever, the British-Dutch conglomerate. A mechanical profile of the potato, which will be recorded as a finite-element computer model, is intended to give Unilever data for genetically engineering the stiffness of cell walls to make tubers more damage resistant.



JOHN McGRAIL

***JULIAN VINCENT** of the Centre for Biomimetics at the University of Reading investigates the fracture properties of coconuts and other natural materials.*

And Vincent believes that understanding the way potato cells change in stiffness by altering their internal pressure might give aircraft designers ideas about changing the shape of wings to control flight aerodynamics. "From a food science, we may find things out that can be of use to the aerospace industry," Vincent says.

The U.S. Air Force, in fact, became interested in Vincent's research. Robert Crane, an air force materials engineer, brainstormed with Vincent recently at Wright-Patterson Air Force Base in Dayton, Ohio, about storing toxic substances in a cellular matrix that can hold large amounts of liquid, a material whose structural properties would be much like those of an apple.

Not everyone thinks Vincent and his colleagues are barking up the right tree. Rustum Roy, a professor of materials science at Pennsylvania State University, insists that little has come of the decades-old endeavor of taking inspiration from biological materials. "This is a very old field that has produced minor advances," Roy says. "What are they all about? They are looking for money."

Biomimetics, in fact, has yet to be incorporated into the mainstream of materials engineering. Westvaco, the New York City–headquartered paper company, adapted ideas from some of Vincent's colleagues in a 1983 patent for a composite in which the orientation of fibers bonded to cardboard is similar to that of cellulose in wood. The tough composite was considered briefly as a candidate for a bulletproof vest. Westvaco, Vincent says, did not quite know what to do with the material, and the project was dropped.

Vincent hopes that a resurgence of interest will win biomimetics growing acceptance as a legitimate subdiscipline within the field of materials research. "You can take ideas from nature and transpose them into technological material," Vincent asserts. "Those ideas are still useful no matter where they come from." *—Gary Stix*

Soul of a New Economic Idea?

The war stories of corporate research laboratories are poignant at best, costly at worst. Remember when Xerox's Palo Alto Research Center held the pieces of the personal computer revolution on its desktops, but the corporation could not see a route to market? Or the time when IBM's John Cocke sketched out ideas for reduced instruction set computer (RISC) designs, but the corporation declined to take up the technology until after others had established tidy businesses?

All too often, the corporate giant that nurtures an innovative idea seems to be the last to make use of it. In many cases, companies are reluctant to undertake the arduous process of bringing a new research idea to market, particularly if it promises to cannibalize an established winner. As a result, upstart firms often topple industrial Goliaths by commercializing radical innovations—even those invented in the labs of the giants themselves.

Such stories run counter to economic theories that paint companies as rational, profit-maximizing organizations. But, curiously enough, even though economists have analyzed decisions as intimate as marriage and divorce, they have spent less time delving into why companies invest in—or eschew—new ideas.

A handful of economists hope to shake up the status quo. Among them, Avinash Dixit of Princeton University and Robert S. Pindyck of the Massachusetts Institute of Technology are trying to codify how a firm decides to invest in, say, facilities for producing a new product. Rebecca Henderson of M.I.T. has been exploring the aspects of large and small firms that might affect the likelihood they will develop—and profit from—radical innovations.

According to the "simple" tale of investment, a firm will expand production or break into a market if the revenue gained by selling the product exceeds the long-run costs of making it. But "reality is very different," Dixit notes. Because companies want their investments to earn some minimum return, they will delay initiating product development until they believe their profits will exceed such a "hurdle."

But what about a firm caught betwixt and between—when existing products are not especially profitable, but future ones fall short of corporate hurdle rates? Pindyck and Dixit suggest that decisions may then be strongly influenced by recent fluctuations in product profitability. "We call it economic hysteresis," Pindyck says.

Physical hysteresis describes how a magnetic field induced in a metal bar by an electric current will linger after the current is shut off. Similarly, the economists suggest, once a firm begins investing, it will continue to do so even if prospective returns slide below the hurdle rate.

Decisions to enter or quit a market, Dixit and Pindyck argue, are in fact driven by the combined effect of three considerations, something economists have typically given short shrift. First, investments carry costs that are either partially or completely irreversible. Second, all such investments are swathed in uncertainty and may never earn the predicted return. And last, companies can control the timing of their investments.

The economists say these characteristics mean that corporate investment decisions can be modeled like decisions to exercise a stock option, a right to buy or sell a security in the future. Companies accrue such options by developing new technologies or by building up their marketing prowess. They exercise these options by exploiting the ideas. A pharmaceutical company that holds "sleeping," or unused, patents essentially has a call option for the lifetime of the patent, Pindyck believes.

Why are big companies typically the last to capitalize on radical innovations?

Viewed as options, corporate decisions not to invest in a novel technology look different. Unexercised options are still valuable, Pindyck points out. Delaying or ignoring an investment opportunity may simply be a rational response to market vicissitudes.

Henderson is less convinced that firms are making optimal choices. But neither is she willing to write off managers as incompetent. Instead she and a handful of other business school professors believe that when firms promote incremental innovation, they inadvertently impair their ability to recognize the value of radical ideas.

Market leaders have long flourished by incrementally improving their products. Such firms enjoy an advantage over newcomers, who lack the intellectual capital and other resources for rapidly improving on the state-of-the-art technology. In the process, these companies put much effort and money into constructing filters to screen out irrelevant ideas. "They can't evaluate all technology instantly, so they learn about the technology they need to understand," Henderson adds.

When faced with proposals for radical shifts, however, those filters become blinders. Such innovations also destroy the intellectual or physical competence that leading firms build by strengthening their existing product lines, Henderson argues. Successful incumbents consequently refuse to take risks that in retrospect would have been rational. Upstarts, which had no particular strength in the older approach, seize the new ideas with vigor.

Henderson bolsters her economic models with evidence from the U.S. semiconductor equipment manufacturing industry. In the late 1970s Perkin-Elmer Corporation was a dominant maker of scanning projection aligners for lithographically etching patterns on silicon. Aligners were soon challenged by an advance called steppers. Although Perkin-Elmer accurately forecasted the progress of the component technologies, the company failed to see how various system elements—such as improved chemicals and lenses—would give steppers an edge. Perkin-Elmer lost its lead to a small firm, GCA.

Henderson says her theories are still in gestation. She is looking for similar trends in other industries. She may have already tripped over one in an area close to home: economics. Established economic theory has long portrayed firms as virtually indistinguishable except for pricing strategies. Henderson's work represents a young economist's somewhat daring explanation of how companies may differ more than her peers believe. "I've been giving seminars to economists," she says, "but it's been running uphill." —*Elizabeth Corcoran*

A Partly True Story

Allow me to introduce myself: Epimenides, professional liar. Well, that's not quite true. My name is really Herman Fenderbender, and I work for a car insurance company. But my friends at the Paradox Club call me Epimenides, and when I'm with them, I always lie.

Last Thursday it was raining, so I got to the club a bit late. Socrates and Plato were leaning against the bar, and next to them was a chubby little fellow.

"This is our newest member, Lukasiewicz," Plato chimed.

"Horrified to meet you," I said in disgust. "My name's Zeno."

"He means he's delighted to meet you, and his name is Epimenides," Socrates explained. "Epimenides always tells lies."

"That's not true," I said. I opened my wallet and took out my business card. "This isn't my card," I commented and handed it over. Lukasiewicz read one side of the card: *The sentence on the other side of this card is true.* He turned the card over and saw: *The sentence on the other side of this card is false.*

"Socrates is right, however. I always tell lies," I boasted.

Lukasiewicz shook my hand warmly. "It's one third false that I'm pleased to meet you, and both sides of your card are half true."

"Pardon?" I said.

"Lukasiewicz is interested in fuzzy logic," Plato explained.

"Instead of just the truth values 1 for a true statement and 0 for a false statement," Lukasiewicz said, "I am prepared to consider half-truths with truth value 0.5 or near-falsehoods with value 0.1—in general, any number between 0 and 1."

"Why would anyone want to do that?" I asked, bemused.

Lukasiewicz smiled. "Suppose I said the club president looks like Charlie Chaplin. Do you think that's true?"

"Of course not!"

"Not even his feet?"

"Well, I guess they do rather—"

"So it's not completely false, either."

"Well, he does look a bit like Chaplin."

Lukasiewicz leaned toward me. He had very penetrating eyes. "How much like him?"

"Around 15 percent I'd say."

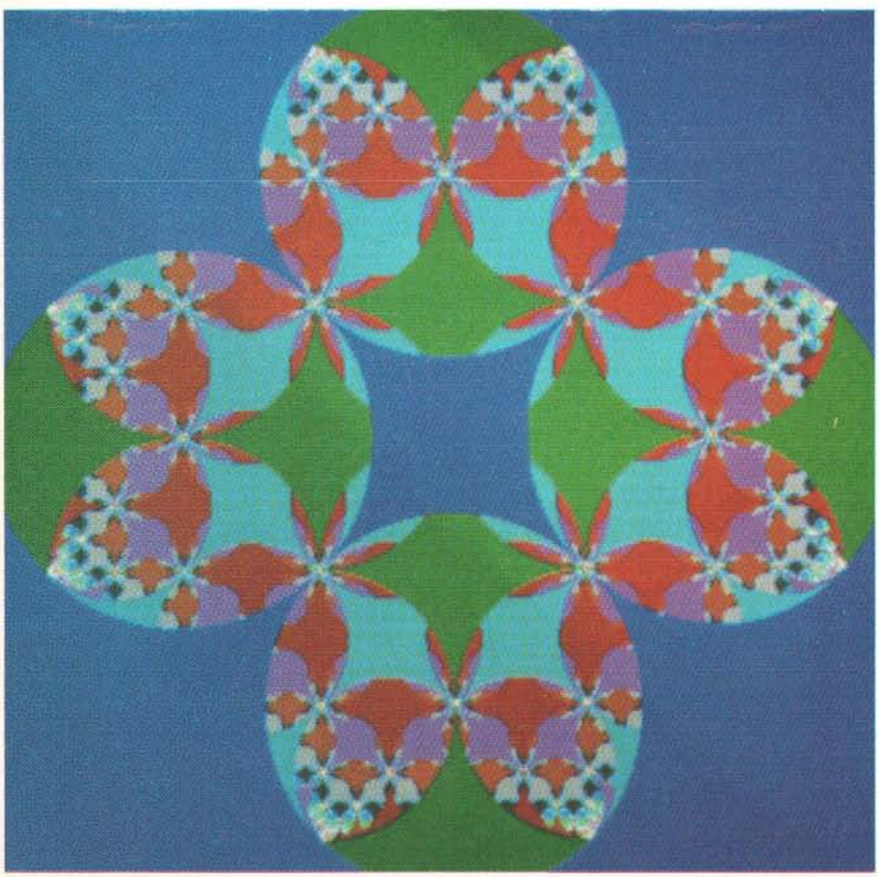
"Good. Then my statement 'the club president looks like Charlie Chaplin' is 15 percent true. It has a truth value of 0.15 in fuzzy logic."

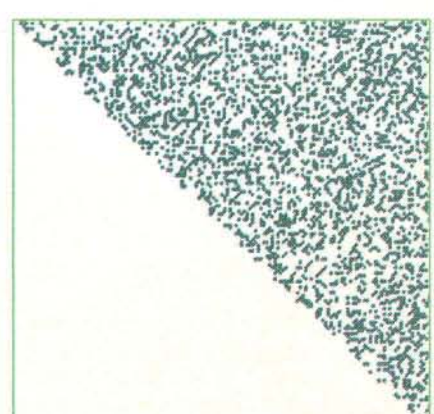
"That's just playing with words. It doesn't mean anything."

Lukasiewicz grasped my arm. "Oh, but it does. It helps to resolve paradoxes. For instance, you claim to be a complete liar. Let's think about your statement 'I am lying.' Or, more simply, 'this statement is false.' In classical logic, it is a paradox, yes? If it is true, then it is false; if it is false, then it is true. To put it another way, you have a statement P with truth value p, which is 0 or 1, and P says the truth value of this sentence is $1 - p$."

"Sorry, I didn't quite get that."

"Ah. If P is true, then its negation, not-P, is false, and its truth value is 0. And conversely. Now, $1 - 0 = 1$ and $1 - 1 = 0$, so if the truth value of P is p, then the truth value of not-P is $1 - p$."





DIAGRAMS OF "ESCAPE TIME"* (left) *and a logical attractor* (right) *were created to analyze a self-referential statement. Such statements typically lead to paradoxes in classical logic. The illustrations above are based on the sentence "the assessed falsehood of this statement is not different from its assessed truth."

"Oh. I see."

"Right. Now the problem is that 'this statement' is P, so P is telling us that the truth value of P is $1 - p$. That's where the paradox comes from. If $p = 0$, then P tells us that $p = 1 - 0 = 1$. And if $p = 1$, then P tells us that $p = 1 - 1 = 0$. Neither choice is consistent."

I gave him a condescending smile. "Luke, all you've done is reformulate in complicated algebraic language what was obvious all along."

He smirked. "Maybe. But in fuzzy logic, there is a consistent solution to the equation $p = 1 - p$, namely, $p = 0.5$. So your claim to be a permanent liar is a half-truth, and everything works out fine. Your own statement leads inevitably to fuzzy logic." Plato slapped him on the back, and Socrates nearly knocked his cocktail over laughing. My face went red, but I saw the point.

"What about his business card?" Plato asked. Lukasiewicz was about to speak, but I stopped him. "Let me answer that. Seems to me I have two statements P and Q with truth values p and q. Moreover, P says Q is true, and Q says P is false. So the corresponding truth-value equations are

$$p = q \text{ for P}$$

$$q = 1 - p \text{ for Q}.$$

Those make no sense if p and q can only be 0 or 1. But there's a unique solution in fuzzy logic: $p = q = 0.5$. So each side of my card is a half-truth, and there's no paradox anymore."

"Precisely," Lukasiewicz said. "But it goes further than that. What we've been discussing is the beginnings of a whole new theory of dynamic logic, invented by Gary Mar and Patrick Grim in the department of philosophy of the State University of New York at Stony Brook. It provides a link between semantic paradoxes and chaos theory."

It was Socrates' turn to look puzzled.

"Oh, wake up. You know what chaos is. Simple deterministic dynamics leading to irregular, random-looking behavior. Butterfly effect. That stuff."

"Of course, I know that," Socrates said in irritation. "No, it was the idea of dynamic logic that was puzzling me. How can logic be dynamic?"

Lukasiewicz looked surprised. "How can it be anything else when discussing self-referential statements? The statement itself forces you to revise your estimate of its truth value. That revised value has to be revised again and again. Consider the Paradox of the Liar, your statement P: 'this statement is false.' Earlier I wrote an equation for its truth value: $p = 1 - p$. But what I should really have written was a process that forces constant revision of your assessment of its truth value, $p \leftarrow 1 - p$. If you assume that P has a particular truth-value p, then P itself tells you to replace that truth value by $1 - p$. For example, if you started out thinking that P was 30 percent true, so that $p = 0.3$, the revision rule implies that $p = 0.7$, which in turn implies that $p = 0.3$ again...and you get an infinite sequence of truth values, oscillating periodically between the two values 0.3 and 0.7. The classical paradox, with $p = 0$ or 1, leads to the sequence 0, 1, 0, 1,...which faithfully reflects the logical argument if P is false, then P is true, so P is false, so P is true, so.... The logical oscillations of the paradox are captured by the dynamics of the truth value."

"And $p = 0.5$ is the only value that doesn't lead to an oscillation," Plato mused.

"Precisely. Now, the Dualist Paradox on your business card is really a logical dynamic:

$$p \leftarrow q$$

$$q \leftarrow 1 - p.$$

Suppose you start out by estimating $p = 0.3$, $q = 0.8$. Then your first revision is to $p = 0.8$, $q = 0.7$. A further revision leads to $p = 0.7$, $q = 0.2$, a third to $p = 0.2$, $q = 0.3$. A fourth revision gives $p = 0.3$, $q = 0.8$, and you're back where you started. It cycles with period four—unless you start at $p = 0.5$, $q = 0.5$, when everything stays unchanged."

"Okay, I'll buy that," I said. "But what about the chaos?"

Lukasiewicz's face went very serious. "Before I can explain that, I must be more precise," he said. "If you want to play around with these ideas for yourself, I'd better tell you how to calculate fuzzy truth values for combinations of logical statements [*see box at right*]. Although all you really need to know at the moment is that not-P has truth value $1 - p$ if P has truth value p. Second, you must know how to assess the truth value of statements about statements."

"I'd like an example," Socrates said.

"Okay. Suppose I said Plato is a good golfer. How true do you think that is?"

"Ooohhh—about 40 percent," Socrates said. Plato gave him a nasty look. "Well, Epimenides usually beats you, and he's pretty mediocre." I gave him an even nastier look.

"Fine. Let's call that statement S. It has a truth value $s = 0.4$. Suppose I make a statement about the statement S. Suppose I utter statement T: 'S is 100 percent true.' How true is statement T?"

Fuzzy Logic

In classical logic, a statement has a truth value of either 1 for true or 0 for false. The statement "the sun is shining" has a truth value of 0 if it is cloudy. In general, statement P has a truth value p equal to 1 or 0. In fuzzy logic, a statement can have a truth value of between 1 and 0. If a cloud obscures a quarter of the sun, then statement P has a value of 0.25.

In fuzzy logic, like the classical theory, the truth value of a statement will change when applying the operators NOT, AND, OR, IMPLIES and IF AND ONLY IF.

NOT-P has a truth value of $1 - p$.
EXAMPLE: If the sun is shining with a truth value of 0.25, then the sun is NOT shining with a truth value of 0.75.

P AND Q has a truth value equal to the lesser of p and q where q is the truth value of statement Q.
EXAMPLE: The sun is shining with a truth value of 0.25, AND Jane is getting tan with a truth value of 0.10.
The value of the example is 0.10.

P OR Q has a truth value equal to the greater of p and q.
EXAMPLE: The sun is shining with a truth value of 0.25, OR Jane is getting tan with a truth value of 0.10.
The value of the example is 0.25.

P IMPLIES Q has a truth value equal to the lesser of 1 and $1 - p + q$.
EXAMPLE: If the sun is shining with a truth value of 0.25, then Jane is getting tan with a truth value of 0.10.
The value of the example is 0.85.

P IF AND ONLY IF Q has a truth value equal to $1 - |p - q|$, that is, one minus the absolute value of p minus q.
EXAMPLE: The sun is shining with a truth value of 0.25 IF AND ONLY IF Jane is getting tan with a truth value of 0.10.
The value of the example is 0.85.

I thought for a moment. "Well, it's certainly not 100 percent true itself. Otherwise S would be 100 percent true, and we've already decided it isn't."

"Right. The degree of truth of my statement T, which is about S, depends on the actual truth value of S and on the truth value attributed to S by T. Here $s = 0.4$, but the value that T leads me to assess is 1. So T will be untrue to the extent that these two values differ, yes? The more inaccurate my assessment, the falser my statement becomes. Because they now differ by 0.6, T is false to the extent 0.6. That is, it is true to the extent 0.4."

"What if you'd said S is half true?"

Lukasiewicz nodded happily. "You'll see how nicely it works. That statement assesses the truth value of S as 0.5, but the actual value is 0.4. The difference is 0.1, which is how false your statement is, so its truth value is 0.9. Because your assessment is only wrong by 10 percent, you're 90 percent correct."

"Ah. And if I'd said S is 40 percent true, I'd have been 100 percent right. So the truth value would have been 1. I've got it."

"Good. In general, suppose I have a statement P with truth value p and a statement Q that leads you to assess the truth value of P to be p'. Then the argument we've just been through says the truth value of Q is $q = 1 - |p - p'|$, where $|x|$ means the absolute value of x (equal to x when x is positive, $-x$ when x is negative). Let me call this the assessment formula."

Lukasiewicz thought for a moment. "Now I can show you what I call the Chaotic Liar, statement C:

This statement is as true as it is assessed to be false.

If its truth value is c, then it instructs you to assess a truth value of $1 - c$. So by the assessment formula, its truth value is $1 - |c - (1 - c)| = 1 - |1 - 2c|$. In short, there is a dynamic process

$$c \leftarrow 1 - |1 - 2c|$$

of reassessment of the truth value c. Choose any starting value for c, say $c = 0.12345$, and calculate successive values. You'll find they are chaotic. Actually, I should warn you that because of round-off errors in your calculator or computer, the process may appear to settle down to either 0 or 1. It may help to replace the dynamic by

$$c \leftarrow 1 - |0.999999 - 2c|.$$

You can even observe the famous butterfly effect of chaos theory—if a butterfly flaps its wings, it can cause a hurricane a month later. More prosaically, small changes in initial conditions make big changes to the subsequent dynamics. If you use a start value of 0.12346 instead, you get a different image."

Lukasiewicz paused. "Next there is the Chaotic Dualist, which involves two statements:

X: X is as true as Y is true

Y: Y is as true as X is false

It's rather like your business card, Epimenides. The dynamics are

$$x \leftarrow 1 - |x - y|$$

$$y \leftarrow 1 - |y - (1 - x)|.$$

To see what it does, you choose an initial pair of values, say, $(x, y) = (0.2, 0.9)$, and calculate successive pairs of values. Think of them as coordinates and plot them in the plane. You get a geometric shape, called the attractor of the dynamic system. In this case, you get a triangle, densely filled with points [*see right illustration on page 85*]. This representation can be transformed into a beautiful and intricate image known as an escape-time diagram. To create it, temporarily relax the conditions that x and y lie between 0 and 1. The idea is to watch how far (x, y) moves from the origin (0,0) and to count how many calculation steps are needed before it goes beyond some threshold value. Then the point (x, y) is plotted in a color that depends on the number of steps required. To start, you should try a threshold value just larger than 1 [*see left illustration on page 85*]."

"I begin to see now," Socrates said. "You take the train of thought involved in assessing the truth value of a set of self-referential statements and convert it into a dynamic process. Then you can apply all the techniques of chaos theory to that process. The escape-time plot is inspired by exactly the same method that creates all those wonderful multicolored images associated with the Mandelbrot set: swirling spirals, sea horses, cacti, stars and so on."

"Indeed. Here's one final idea for you to mull over. We can rephrase the Chaotic Liar as

The assessed falsehood of this statement is not different from its assessed truth.

In fuzzy logic, it is standard to interpret the adjective 'very' as forming the square of a truth value. So think about the rather woollier statement

The assessed falsehood of this statement is not very different from its assessed truth.

The statement leads to the dynamic

$$p \leftarrow 1 - (p - (1 - p))^2,$$

which converts into the form

$$p \leftarrow 4p(1 - p).$$

Chaos theorists call this the logistic dynamic system—so my statement is the Logistic Liar. It's chaotic, too—try it."

At midnight, the Paradox Club closed, and Lukasiewicz and I walked out into the street. I realized I had been so absorbed working out examples of fuzzy-logical chaos that I had forgotten to ask one very important question. "Luke, it's all very pretty, but how significant is it?"

"Well," he said. "Mar and Grim point out that it gives a geometric approach to semantic complexity, letting you distinguish between different systems of self-referential statements. They also say it can be used to prove there is no decision procedure that will tell you whether or not a given system is chaotic. That's a result in the same general line as Kurt Gödel's famous theorem on the undecidability of arithmetic. It's potentially rather deep stuff, Epimenides."

"So I see. Connections between logic and chaos— Amazing! But wait a second. How can I be sure everything you just told me is true?"

"If I have ever lied to you, I ask the gods to strike me down with two lightning bolts."

Just then, thunderclouds formed in the sky, and a single lightning bolt zapped Lukasiewicz into oblivion. I looked up, shaking my fist at the clouds: "So was he telling me the whole truth, or only half?"

FURTHER READING

DOES GOD PLAY DICE? THE MATHEMATICS OF CHAOS. Ian Stewart. Basil Blackwell, 1990.

PATTERN AND CHAOS: NEW IMAGES IN THE SEMANTICS OF PARADOX. Gary Mar and Patrick Grim in *Noûs*, Vol. 25, No. 5, pages 659–693; December 1991.

SELF-REFERENCE AND CHAOS IN FUZZY LOGIC: RESEARCH REPORT #92-01. Patrick Grim. Group for Logic and Formal Semantics, Department of Philosophy, S.U.N.Y. at Stony Brook, 1992.

COMPUTER INVESTIGATIONS IN THE SEMANTICS OF PARADOX: CHAOTIC LIARS, FRACTALS, AND STRANGE ATTRACTORS. Gary Mar and Patrick Grim in *Philosophy and Computing* (in press).

BOOK REVIEW by Nathan Keyfitz

The Preservation of the Planet

The time has come when scientific truth must cease to be the property of the few, when it must be woven into the common life of the world.

—Louis Agassiz, 1863

ONLY ONE WORLD: OUR OWN TO MAKE AND TO KEEP, by Gerard Piel. W. H. Freeman and Company, 1992 ($21.95).

The last third of the 20th century has been marked by an issue rarely considered prior to the generation of our parents: the preservation of the planet. Astronomy and space exploration bring reports from the outside world that confirm what biology and geophysics working from the inside tell us emphatically. The planetary mechanisms governing climate, soils and plants are complex, fragile and probably unique in the universe—and we depend on them totally. The awareness that our lives depend on mechanisms we are far from understanding is unnerving. The more so when we realize that industrial progress has increased the scale of human activities to the point at which they can be counted a major terrestrial force, comparable in magnitude with the forces of nature.

A matter so important has stimulated much print, and this may be the thousandth book on it, but no other is more deeply immersed in the achievements of science and in whatever is known about the damage done to our environment by science-based technology. It examines the achievements from a thoroughly humane viewpoint and never ceases to be nonpartisan. The author stands between the embattled armies of the proponents of growth and consumption on the one side and, on the other, those who want the least possible change from the pristine planet that existed before our distant forebears came on the scene. It goes as far as any one book can go toward putting into effect Agassiz's admonition.

Gerard Piel is both a devoted statesman of science and a critic of many of the uses made of it. One of the founders of *Scientific American* in the form in which we know it, he has spent his days for half a century in the labor of scientific dissemination, and he submits his accumulated knowledge to the test of discerning where the world is heading. He shows little trace of the thoughtless arrogance so often associated with technical achievement, an arrogance that is an even greater hazard for laypeople who have identified themselves with science than for scientists themselves.

One of the achievements of the June 1992 United Nations Conference on Environment and Development (UNCED) in Rio de Janeiro was to link permanently in our thought economic development and protection of the environment. This book presents the nature of that linkage. In Piel's words:

> Local air pollution, regional acid rainfall, apparent global warming, and the "hole" in the ozone layer all testify to the planetary reach of the industrial revolution. Taken together with noxious solid-waste landfills, toxic chemical dumps, the acidification of lakes, the contamination of groundwater, streams, and estuaries, the erosion of the soil on all continents, and the impoverishment of ecosystems invaded by human habitation, these offenses of the present and threats to the future have put "the environment" on the political agenda of every country in the world.
>
> All of this is the price the world has been paying for the good fortune of the 25 percent of the population that lives in the industrial countries. They consume 75 percent of the now so vastly expanded world output of material wealth. For the first time, whole populations of nations are rich.

The massive impact of humankind on the planet did not start with modern industry. Long before the industrial countries came on the scene, unsustainable types of agriculture were in use, and these continue into the present: slash-and-burn practices account for 35 percent of the deforestation in the Amazon region, 50 percent in South and Southeast Asia and 70 percent in Africa.

But in earlier epochs, populations had been small enough that they could carry on slash and burn in reasonable symbiosis with the forest, cultivating land for only two or three years before moving on. Forty or 50 years later the forest had substantially recovered, and as long as the population was small enough in relation to the land at its disposal, so that it did not have to return too early, this long fallow cultivation could continue indefinitely. It is population growth that disturbs the stable relation with the forest. Maurice Strong in the foreword is broader and more emphatic: "The quality of future life on Earth will depend to a high degree on the size of the stable world population that is likely to be reached some time in the next century."

This destabilizing effect of population has in some societies been stimulating. "People have doubled the food supply...to meet...doublings of their numbers." Where possible they did so by doubling the area under cultivation. Where that was impossible, they were forced to move up the scale of technical and economic progress. Some enthusiasts have claimed as a general rule that progress results from overpopulation, but going around the world counting cases would show an immense majority of the overcrowded lands in which population growth has reinforced a perennial condition of want.

We can never have exact data on the age-old condition, but the recent past exemplifies in statistics this income immobility of the poorest countries. Bangladesh showed an income in 1970 of U.S. $170 per capita; in 1990, after some ups and downs, it had returned to exactly the same figure. Its GNP had gone up by 70.4 percent, but this was offset by the rise in its population from 66.7 million to 113.7 million, just the same percentage. Could Bangladesh have had the same increase in its GNP if its population had been level? Some would say yes, that because of land shortage and lack of capital the marginal population brings little to the income of the community. But after all, others say, following John Stuart Mill, every mouth comes with a pair of hands, and surely people in their desperation will think of some way of gaining a livelihood, if only they are not hampered by government. That way of seeing the matter is more congenial to us in the advanced countries than to those closer to the action. The Third World, as innumerable surveys have shown, has an unsatisfied desire for information and equipment to control births. According to the World Fertility Survey, some 25 to 50 percent of Third World women of fertile age want the wider choice that command of birth control would give them.

But there is a good side to this explosion of population; Piel goes so far as to call it a "benign event." The increase

in population is not caused primarily by increasing births but by "the lengthening of their life expectancy...the beginning of their industrial revolution." Food shortage was the result of excessive population growth most feared through the 19th century, yet a member of the increased population of today typically disposes of more food than his or her ancestors did. In no area of endeavor has the achievement of science been more impressive than in agriculture, which has banned for most of the next century, perhaps forever, the specter of humankind's inability to feed itself. The starvation that provides a tragic spectacle to television viewers is attributable to causes quite other than inability to raise enough food in total.

The story of China that Piel relates offers an example of the importance of population control. Once it had overcome the idiosyncratic view of population promoted by Mao and inherited from Marx, China moved effectively to hold down its population increase. China's total fertility rate fell from 6.24 in 1950–56 to 2.30 in 1985–90, compared with the less developed countries as a whole at 6.19 to 3.90 over the same period. "With a net reproduction rate not far above 1, China is approaching zero population growth at a lower standard of living than any nation that has preceded it into industrial revolution." The relation of this accomplishment to literacy, especially the education of women, and to confidence that somehow the old age of parents would be secure, needs little underlining. "The sense of economic equity abroad among the people...has made it possible for the national government to promote population control by draconian measures," including the rationing of pregnancies and the decree of one-child families.

Piel discusses with similar detail the way development is going in other areas of the world, but the Chinese case has been especially revealing. Comparison with the U.S.S.R. and Eastern Europe shows that only the name of communism was shared. Under it, countries have been ruled by ruthless dictatorships, more than ready to suppress alien peoples, but aside from that unpleasant common feature the rule has been as varied as are national cultures and temperaments.

Now that all are trying to escape into free markets, China has an advantage. It freed its farmers to sell their own produce in the early 1980s and allowed foreign-operated and foreign-controlled industry almost as early. That has meant a much less sudden break with the past than Russia and Eastern Europe are having to face. The same pragmatism appears in the way that markets are introduced: first in sectors of the economy and in regions of the country where the infrastructure is adequate and the capital required is small. These practices permitted an immediate acceleration of production under private ownership, especially in South China and the area around Guangdong (Canton).

Looking at the other side of the world, what is called the North, Alfred Sauvy's "First World," Piel has a sharp eye for the difficulties of an economic system that has matured—in the sense that material want, which has tried the human race from the beginning, no longer exists for the population as a whole in the advanced countries:

> They are rich in the fundamental sense that, for them, the economic problem defined by Adam's first question after the Fall—'When do we eat?'—is solved. For them, toil is no longer their living and their life. Their work is done by machines in progressively workless economies. In the U.S. economy, the climax to which all are trending, less than 3 percent of the so-called labor force remains engaged in agriculture; less than 30 percent in the production of goods.

At per capita incomes of $16,000 in current dollars in the European Community, $21,000 in the U.S., $24,000 in Japan, $30,000 in Switzerland, no one need suffer hardship. Our problems arise from quite other causes than shortage of basic goods.

Our economic and social institutions have failed to adapt to the technology that has made physical work, the kind of work that prevailed from the beginning up to well into this century, entirely obsolete. Not only has arduous physical work disappeared in the advanced countries, but even light work that is repetitive is now taken over by automation. And repetitive mental work is following that into obsolescence as computers dominate the workplace. Much of the repetitive work that remains is the keying of words and numbers from written or printed text, and most of this will be eliminated as soon as scanning apparatus is made more reliable.

Among many other signs of the lack of adaptation is persistent unemployment, on the one side, and on the other side purposelessness. Our ancestors, whose work was hard and often dangerous, always necessary simply to keep alive, seemed to know what they were here for. Now "anomie and preoccupation with the isolated self recur as a central theme of U.S. popular culture. That they find resonance in every other industrial country suggests that the solving of the economic problem brings on these quandaries everywhere. The answer to the simplifying question of survival asks a host of new questions for which there are no ready answers."

Attempts at explanation of this maladaptation have many traps into which the explainer can fall. No one can doubt that the dominance of the question of sheer survival in all past times simplified the life philosophy of those who had to face it. Yet why does its solution release all the commotion that we know, the personal as well as the social disorganization? Piel cites one answer by a distinguished social psychologist, in which the blame is placed on the two-child family, "an inadequate institution for the socialization of the young." Children are "in minimal contact...with adults, including their own parents."

But, I say, that is surprising; one would think that with the small family, children would have far more opportunity for contact with adults, especially their parents. If they do not, then it is not the small family that is to be held responsible. On the contrary, it must rather be some force that overcomes the smallness of the family. It must be some unidentified change in the underlying culture. If that change had not occurred, the contact with parents would surely be greater with two children per family than with eight. I do not pretend that my expression "change in culture" constitutes a satisfying explanation; it is closer rather to an assertion that we do not know the explanation. My point here is only that admitting ignorance is better than attribution of lack of contact with adults to the small family—that is, to something that cannot be its cause.

It is especially in the past few years that there has been a lag on the economic part of adaptation:

> Technological disemployment has been absorbed up to now in the United States and other industrial countries, by resilient economic and social institutions. Shorter and fewer working days in the year and years of prolonged education and earlier retirement may be counted the principal product of automatic production. In the United States, nearly half of the rising generation goes on to some higher education; such opportunity is opening to the young throughout the industrial world. The value of this time eludes quantification; it is not reckoned in the gross national product. The values beyond economic value to be had from the exercise of human capacity liberated from the compulsions of getting a living defy imaginations presently bound to that task.

One failure of institutions is reflected in distribution, on which, as Piel says, economics has never had much to say. It concerns itself with efficiency and allocation, where it has remarkable achievements, but rarely discusses allocation between rich and poor. And the failure to handle that is one of the causes of the problems that the advanced societies face. These problems are no less real for being unmeasurable in dollar terms: among the young and the poor, crime, drugs and suicide; among the better off, confusion of values. Projecting present tendencies, one can readily imagine increasing prominence of inequity between rich and poor, of restless subnational societies, each with its own culture demanding recognition, of fundamentalist religious communities gathering into their folds both those that economic advance has bypassed and those who find too little satisfaction in what economic advance does for them. And these sources of social instability are set in a physical world that is itself unstable as our system of production and consumption multiplies assaults on the environment. Now that the economic questions are partly behind us, perhaps the social sciences, including economics, which is the most advanced among them, could turn their attention in this direction more wholeheartedly than they have so far done.

This debasing of the natural environment in which we and our economy have to sit is a charge on the future. Only so much forest can be cut worldwide because there is only so much there. Only so much greenhouse gas can be released without intolerable results. What we draw of the tolerance of the biosphere, our children will not be able to draw.

Very well, some say, that is an economic problem; people should not be allowed to cut trees or release greenhouse gas—in general, to carry on activities for which others are being forced to pay part of the cost—and then to retain the whole profit. If people driving less will preserve the environment, then a suitable tax on gasoline will push them in that direction. Once they cover all the costs of driving, the market will ensure that just the right amount of driving is done. Gasoline in Europe is taxed at 75 U.S. cents per liter; if necessary to secure the desired reduction in the U.S., then such a tax should be imposed. We have here a purely economic problem to which there is an economic solution, discussed in textbooks under the heading of externalities.

Not for Piel is it a purely economic problem. He puts reducing contamination of the environment, along with good schools and maintenance of the infrastructure, in the category that "require[s] political as well as market choices." And once we come to politics we must not think that politicians will put some measure into the laws just because they are advised by experts to do so. When the U.S. administration proposed a seven-cent (not 75-cent) federal tax on gasoline the Congress rejected it. Congress will vote laws that its constituents urge it to vote. It is curious that advisers who firmly believe in market economics—which holds that when people act according to their personal interests the social product is maximized—urge legislators to do something (tax gasoline, balance the budget) that will lose them their seats. Without public understanding, there will be no action.

The academy, of course, helps in the education of the public, but it is held back by an innate conservatism. That conservatism shows itself in unwillingness to recognize fundamental social change. The unwillingness comes out most conspicuously in the powerful methodologies being developed to solve problems that history has already solved for us, or simply bypassed.

Indeed, academics are not mere eccentrics writing purely out of their heads; they partake of the public consciousness of their time. It is an odd circumstance that in this moment when the economic problem, insofar as it is the aggregate production of the goods needed to sustain life, is essentially solved for the advanced countries, academics as well as the public are more preoccupied by it than ever. Ever since World War II no measure has been more commonly referred to than the GNP. Its rate of current increase is everywhere taken as the most precise index of the happiness of nations. While we stand hypnotized by that aggregate measure we could well come to grief through failure to recognize the deadly threats hiding in four quite different needs that the rising aggregate income seemingly does not help to satisfy: First, the division of the economic product within the rich countries, and most especially providing opportunities for employment among all those able and willing to participate in production. Second, the development of the poor countries. Third, the integrity of the global environment. And last, the defense of social order against attack by crime and the alienation of the young.

The Enlightenment produced two giant figures, Adam Smith and Karl Marx. Each carried conviction in his time. Smith's prescriptions spread and are still spreading, generating luxury and relief from physical toil for the mass-

es. Marx's prescriptions gave the masses little but poverty and slavery in those countries that professed to follow them. What Smith and Marx have in common is that the doctrines of both are historically bounded. There is no reason why either should prevail for the indefinite future.

None of the four groups of needs listed above is being met by the calculus of the market. I read in this book what I read in John Maynard Keynes. Keynes spoke of that future day, which in 1931 he thought only half a century away, when the economic problem would have been solved for most citizens of Europe and America: "The love of money as a possession—as distinguished from the love of money as a means to the enjoyments and realities of life—will be recognized for what it is, a somewhat disgusting morbidity, one of those semi-criminal, semi-pathological propensities which one hands over with a shudder to the specialists in mental disease."

Keynes goes on to express the hope that greed, a temporary necessity in the period when people had to be motivated to work and abstinence, to accumulate capital in the face of scarcity and hardship, will give way to a return to the classical virtues of altruism and sacrifice of self, as they appear in the recorded literature of all ages before the industrial revolution. With the facilities we now have, we should be able to come closer to practicing these virtues than did our ancestors who often merely preached them.

Piel's exposition implies at many points that those ancient and universal virtues fit with the present predicament of the rich countries as the virtue of self-seeking fitted with the harsh requirements of the industrial revolution. He and Keynes are not alone in this view. A significant minority of the young people of America and Europe are feeling their way to an escape from the ethic of greed: eating less meat, dressing simply, backpacking to see the world, looking for ideas in the primordial religions of Asia. They point to a more hopeful future than the still dominant debt-financed, planet-destroying, ostentatious living of us, their parents. What gives *Only One World* vitality is its combining the healthy dissatisfactions of the young with a mature appreciation of the power of science and the industrial economy to move us to a more sustainable course than the one we have been following.

NATHAN KEYFITZ, Andelot Professor Emeritus at Harvard University, is senior scholar at the International Institute of Applied Systems Analysis in Austria.

Germany and the Bomb: New Evidence

Why didn't Germany build an atomic bomb during World War II? This question has yielded divergent answers over the past 48 years. After the war, German scientists were unwilling to admit any support of German war aims but were eager to bring their nation into the nuclear age with a nuclear power program under their own direction. They therefore argued that their failure was not a scientific one but was the result of wartime shortages and their own lack of will. Some even hinted at a moral rejection of building bombs for Hitler. Allied scientists, in contrast, had to confront the unwanted results of their achievement—the onset of a nightmare era of nuclear cold war and military control of research. They consoled themselves that they had done their work in what they believed to be a race against an effort to put a nuclear weapon in Hitler's hands: if the Germans had failed, they argued, it was not for lack of trying.

The long-awaited release to the public in early 1992 of secretly recorded conversations among 10 German atomic scientists has inspired yet another round of divergence. The scientists were held in Allied captivity after World War II at an English country estate, Farm Hall, near Cambridge, and their conversations were recorded by microphones hidden in the rooms at Farm Hall. Among the 10 were Otto Hahn, codiscoverer of fission, Walther Gerlach, physicist and administrator of the German atomic project, and Werner Heisenberg, scientific head of fission research.

The newly released conversations indicate that the scientists' account originated in the shocked aftermath of the news that the U.S. had dropped an atomic bomb on Hiroshima on August 6, 1945. The captive Germans had thought themselves ahead of the Allies all along. Realizing that the Allies did have the bomb, their conversation turned to two main issues: How did it work? and Why had Germany not achieved a nuclear weapon? After a night of agonizing by all, Heisenberg and Gerlach prepared a press release articulating the scientists' official position.

The documents recently released are not the original tapes or even transcripts of them. They are biweekly intelligence reports to Washington, containing idiomatic British translations of excerpts of the conversations. The 270 pages of text had long been available to several key intelligence figures, most notably Samuel A. Goudsmit, who used them in his 1947 account of Allied intelligence about the German atomic effort, *Alsos.* Nor are the newly public reports the only materials available from Farm Hall. Over the years, a diary, numerous letters and the press release written by the scientists while at Farm Hall have come to light. What is different is that, for the first time, the *public* has access to a record greatly enriched by the scientists' unguarded conversations. Against the backdrop of the wartime record, the material just released enables a clearer insight into the difficult question of the German failure and into the origins and ultimate inaccuracy of the German scientists' postwar account.

Two sharply differing interpretations of the Farm Hall debates have recently appeared in the U.S. Stanley Goldberg and Thomas Powers, writing in the *Bulletin of the Atomic Scientists,* support the German position as it evolved at Farm Hall and even go several steps further: Heisenberg delayed the effort, hid their progress from the authorities and leaked information to the Allies. Jeremy Bernstein, writing in the *New York Review of Books,* finds general support for Goudsmit's earlier assessment: the German scientists did not know how an atomic bomb works and raised morality as a foil for their failure.

One potentially decisive question has always been: How much of the fissionable but rare isotope uranium 235 did Heisenberg think would be needed for a bomb? If it were a relatively small amount (which it is), they could have obtained it if they had really tried; a large amount would undercut the need for scruple. "Quite honestly," Heisenberg tells Hahn at Farm Hall, "I have never worked it out, as I never believed one could get pure 235." But Heisenberg did know that a reactor would breed the easily extracted plutonium, which could also power a bomb—leaving the question of scruple here unresolved.

While Bernstein faults Heisenberg on other facets of bomb research, Goldberg and Powers press their case for sabotage. Their main evidence beyond Farm Hall is a newly found thirdhand intelligence report reaching Washington in April 1941. It stated that Germany was working on a bomb under Heisenberg but that Heisenberg delayed the work, fearing "the catastrophic results." There is, however, no evidence of any delay at that time. In fact, in a February 1942 meeting with regime officials, Heisenberg readily explained that a reactor could be built, that an atomic explosive was possible and that the reactor would breed plutonium.

Yet Heisenberg *was* reticent in a June 1942 meeting with Albert Speer, Hitler's head of wartime industrial production. Reports of the meeting suggest that Heisenberg did not mention plutonium as a reactor by-product. Speer was satisfied enough with the prospect of a novel energy-producing device to continue modest support, while concentrating funds on the development of rockets and jet aircraft.

Why were Heisenberg and his colleagues suddenly less forthcoming? In my opinion, the answer entails neither scruple nor stupidity, but technology and circumstance. Speer was gearing the German economy for total war. If the scientists had hinted that they could produce a bomb, the regime would have established a crash program. Failure to produce in short order (a likely outcome) would have had severe consequences. Believing themselves far ahead, the physicists felt that an explicit bomb program was unnecessary—yet within their grasp if they so desired it. Moreover, by encouraging only modest reactor support, the German scientists received everything they wanted: financing, recognition, control over the project, protection from ideological attack and reduced risk of assignment to the front coupled with an opportunity to support the German cause. War conditions soon made their decision irreversible.

In the shadow of defeat, the Nazi atrocities and the Allied achievement, the Farm Hall scientists suddenly found the reasons just outlined to be a politically untenable explanation for their wartime failures. They responded with an account of the past tailored to their present situation: moral scruple compounded with administrative and material limitations far beyond their control.

DAVID C. CASSIDY is an associate professor at Hofstra University. His book Uncertainty: The Life and Science of Werner Heisenberg *was recently released by W. H. Freeman and Company.*